CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
PROFMAT – Mestrado Profissional em Matemática em Rede Nacional

ODILON ANTÔNIO BORGES GOMES

ESTRATÉGIAS PARA O ENSINO DA ESTATÍSTICA NA EDUCAÇÃO BÁSICA

Belo Horizonte
2020

ODILON ANTÔNIO BORGES GOMES

ESTRATÉGIAS PARA O ENSINO DA ESTATÍSTICA NA EDUCAÇÃO BÁSICA

Dissertação apresentada ao Centro Federal de Educação Tecnológica de Minas Gerais como parte das exigências do Programa de Pós-Graduação Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, para obter o título de Mestre.

Orientador (a):
Marcela Richele Ferreira

Coorientador (a):
Ricardo Saldanha de Morais

Banca Examinadora:
Dênis Emanuel da Costa Vargas
Fernanda Aparecida Ferreira
Lillia dos Santos Barsante Silva
Stella Maris Lemos Nunes

Belo Horizonte
2020

Gomes, Odilon Antônio Borges
G633e Estratégias para o ensino da estatística na educação básica / Odilon Antônio Borges Gomes. – 2020.
118 f.

Dissertação de mestrado apresentada ao Programa de Mestrado Profissional em Matemática em Rede Nacional.
Orientadora: Marcela Richele Ferreira.
Coorientador: Ricardo Saldanha de Morais.
Dissertação (mestrado) – Centro Federal de Educação Tecnológica de Minas Gerais.

1. Estatística – Estudo e ensino – Teses. 2. Modelos matemáticos – Teses. 3. Aprendizagem – Teses. 4. Ensino fundamental – Teses. I. Ferreira, Marcela Richele.II. Morais, Ricardo Saldanha de. III. Centro Federal de Educação Tecnológica de Minas Gerais. IV. Título.

CDD 519.5

Elaboração da ficha catalográfica pela bibliotecária Jane Marangon Duarte, CRB 6º 1592 / Cefet / MG

ODILON ANTÔNIO BORGES GOMES

ESTRATÉGIAS PARA O ENSINO DA ESTATÍSTICA NA EDUCAÇÃO BÁSICA

Dissertação apresentada ao Centro Federal de Educação Tecnológica de Minas Gerais como parte das exigências do Programa de Pós-Graduação Mestrado Profissional em Matemática em Rede Nacional – PROFMAT, para obter o título de Mestre.
Orientador (a): Marcela Richele Ferreira
Coorientador (a): Ricardo Saldanha de Morais

APROVADA: 26 de novembro de 2020.

Odilon Antônio Borges Gomes

Ricardo Saldanha de Morais
(Coorientador)

Marcela Richele Ferreira
(Orientadora)

Belo Horizonte
2020

AGRADECIMENTOS

Começo agradecendo a Deus, que ilumina meus caminhos e me ampara nos momentos difíceis, dando-me força e determinação para concluir mais uma importante etapa de minha trajetória.

Um enorme agradecimento a meu pai Odilon, *in memorian*, e a minha mãe Josefina que sempre colocaram como o principal legado para seus filhos o amor, a educação e a fé!

Agradeço, com profundo carinho, a toda minha família pelas palavras de apoio, de incentivo e de afago quando percebiam meu cansaço na jornada de estudos e trabalho. Ao meu irmão Fernando, que tanto me inspira na carreira acadêmica. A todos que sempre suportaram e compreenderam minhas ausências para que eu conseguisse concluir esta etapa. Em especial, agradeço a minha esposa Débora e minhas filhas, Letícia e Alice, pois sem elas não teria conseguido. Amo vocês! E aqui, não posso deixar de mencionar o quanto me fazia bem as mensagens do meu sogro, Ilson Lessa, “força e boa semana”!

Agradeço a minha orientadora Marcela e meu coorientador Ricardo, sempre comprometidos em me ajudar e incentivar ao longo deste trabalho. Aos professores, secretários do Departamento de Matemática do Cefet-MG e ao secretário do programa Profmat Pedro Falci Cardoso, que não mediram esforços para que o programa obtivesse o sucesso devido. Aos membros da banca examinadora, pelo tempo dedicado a este trabalho e as sugestões apontadas.

Aos meus colegas do Mestrado, principalmente o Homero, com quem tive o prazer de trocar experiências ao longo de todo o curso. Enfim, a todos que fizeram parte direta ou indiretamente desta caminhada!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

Dados do Ministério da Educação, extraídos do exame do PISA em 2018, apontam que os educandos brasileiros têm baixa proficiência quando são chamados a resolver problemas pelas ferramentas matemáticas. Os profissionais da educação, que se dedicam ao ensino da matemática, devem estar comprometidos com a formação de seus alunos exercitando atividades que trabalhem a investigação, a reflexão e a análise crítica para a resolução de problemas atrelados às demandas cotidianas das atuais e futuras gerações de profissionais e cidadãos. É preciso preparar o educando para atender às exigências de uma realidade globalizada cujas informações, claras e baseadas em dados fidedignos, são formas oportunas de disseminar os conhecimentos. Fundamentado em uma pesquisa bibliográfica e proposições de modelos matemáticos, esse trabalho propõe uma abordagem alternativa para o ensino da Estatística Descritiva e da Probabilidade, visando auxiliar professores da Educação Básica a desenvolver estratégias de ensino por contextualização, investigação e interdisciplinaridade, utilizando como recurso um banco de dados de situação real. Buscou-se selecionar, investigar e descrever estatisticamente dados relevantes como o censo demográfico, imunização contra a doença do sarampo, os casos registrados por infecção pelos vírus do sarampo e da dengue, para utilizar a técnica de modelagem matemática, baseada na metodologia da sequência didática, como ferramenta auxiliar de ensino. Para cada um destes quatro bancos de dados modelados, elaborou-se análises sequenciais que trabalham a compreensão e a posterior descrição do conjunto de dados por meio de gráficos, tabelas e argumentações embasadas por cálculo probabilístico. As construções citadas foram elaboradas no *software* Excel por ser uma ferramenta de fácil manipulação pelo usuário. Porém, toda a atividade também pode ser implementada em um *software* livre como o Openoffice, Libreoffice, *Software* R, dentre outros. A finalidade é proporcionar ao educando uma melhor compreensão da Ciência que trata da análise e interpretação das informações, de forma a promover a aprendizagem significativa das definições da estatística e da teoria da probabilidade. A proposta de atividade não foi implementada devido a Pandemia da Covid 19 e a consequente interrupção das aulas presenciais. Todavia, desenvolve-se um estudo que torna possível estabelecer uma metodologia que permite ao educando transformar um grande volume de dados em um material organizado, informativo e útil para análises críticas sob diferentes aspectos científicos, tecnológicos e sociais.

Palavras-chave: Estatística Descritiva. Modelagem Matemática. Aprendizagem Significativa.

ABSTRACT

Data from the Ministry of Education, according to PISA exam in 2018, point out that Brazilian students have low proficiency when asked to solve problems by using mathematical tools. Education professionals, who are dedicated to teaching mathematical tools, must be committed to train their students by practicing activities that work with research, thinking and critical analyses to solve problems linked to the daily demands of current and future generations of the professionals and citizens. It is necessary to prepare the learner to meet the requirements of a globalized reality whose clear information, based on reliable data, are opportune ways to knowledge dissemination. Based on a bibliographic research, propositions of mathematical models and by using a data analysis software widely accessible to society, this work proposes an alternative approach to the teaching of Descriptive Statistics and Probability, aiming to help Basic Education teachers to develop teaching strategies by contextualization, investigation and interdisciplinarity, using a real situation database as a source. This work aimed to select, investigate, and describe statistically relevant data such as demographic census, immunization against measles, cases registered of measles and dengue infections, to use the mathematical modeling technique, based on the methodology of the didactic sequence as an auxiliary teaching tool. For each of these four modeled databases, sequential analyses were developed for improving the understanding and subsequently description of the data set by means of graphs, tables, and arguments based on probabilistic calculations. The constructions mentioned were elaborated by using Excel *software*, since it is a tool of easy manipulation by the user. However, all activity can also be implemented in free *software* such as *Openoffice*, *Libreoffice*, *Software R*, among others. The purpose is to provide the student with better understanding of Science which deals with the analysis and interpretation of information in order to promote meaningful learning of the definitions of statistics and the probability theory. The proposal activity was not implemented due to the Covid-19 Pandemic and the consequent interruption of face-to-face classes. Nevertheless, it becomes a study that makes possible to establish a methodology that allows the user to transform a large volume of data into organized, informative, and useful material for critical analyses under different scientific, technological, and social aspects.

Keywords: Descriptive Statistics. Mathematical Modeling. Meaningful Learning.

LISTA DE SÍMBOLOS

$\approx$ Aproximadamente

$\overline{A}$ Complementar de um conjunto

B^c Complementar de um conjunto em relação ao conjunto universo

{ } Conjunto

$\emptyset$ Conjunto vazio

$\neq$ Diferente

Ω Espaço amostral

$\subset$ Está contido

$\cap$ Interseção

$\vdash$ Intervalo fechado à esquerda e aberto à direita

$>$ Maior

$\bar{x}$ Média

$\tilde{x}$ Mediana

$<$ Menor

$\leq$ Menor ou igual

x^* Moda

$\not\subset$ Não está contido

$\in$ Pertence

Σ somatório

$\cup$ União

μ Valor esperado

LISTA DE ABREVIATURAS

BNCC – Base Nacional Curricular Comum

BNCC-EM – Base Nacional Curricular Comum - Ensino Médio

Cebraspe – Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos

CEFET/MG – Centro Federal de Educação Tecnológica de Minas Gerais

CNE – Conselho Nacional de Estatística

CNE – Conselho Nacional de Educação

CNG – Conselho Nacional de Geografia

CNE/CP – Conselho Nacional de Educação/Conselho Pleno

dm – Desvio médio

dp – Desvio padrão

EPMEM – Encontro Paranaense de Modelagem na Educação Matemática

Fiocruz – Fundação Oswaldo Cruz

IBGE – Instituto Brasileiro de Geografia e Estatística

IBGE-SIDRA – Sistema IBGE de Recuperação Automática

INEP – Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira

log – Logaritmo

OCDE – Organização para Cooperação e Desenvolvimento Econômico

OMS – Organização Mundial da Saúde

PISA – Programa Internacional de Avaliação de Estudantes

PNE – Plano Nacional de Educação

PROFMAT – Mestrado Profissional em Matemática em Rede

SIH/SUS – Sistema de Informações Hospitalares do SUS

SUS – Sistema Único de Saúde

Trim. – Trimestre

var – Variância

LISTA DE GRÁFICOS

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

Muito se discute sobre como melhorar a qualidade da Educação Básica no Brasil. Nesse sentido, um passo fundamental é usar de maneira eficiente o tempo que o educando permanece na escola. Ademais, como a aprendizagem é um processo contínuo e individual, próprio de cada sujeito, conduzir sabiamente esse tempo com atividades que confiram sentido aos conteúdos curriculares ensinados, é uma forma de persuadir o aluno levando-o a se interessar pela aula. Por fim, a expectativa é que esse interesse favoreça a aprendizagem significativa dos temas abordados.

Segundo Ausubel (2003), a aprendizagem significativa ocorre quando o indivíduo abstrai o conhecimento que lhe é proposto, agregando mais ao saber cognitivo e conhecimento prévio. Assim, compreender e relacionar os fenômenos estudados contribui para a sua contínua formação, dando-lhe capacidade para analisar variáveis socioeconômicas e técnico-científicas. Em síntese, quando a evolução da capacidade de compreender, relacionar e analisar é interrompida, percebe-se que alguma parte da aprendizagem ficou sem sentido (aprendizagem mecânica). Moreira (2012, p. 04) coloca que

> [...] a aprendizagem significativa não é, como se possa pensar, aquela que o indivíduo nunca esquece. A assimilação obliteradora é uma continuidade natural da aprendizagem significativa, porém, não é um esquecimento total. É uma perda de discriminabilidade, de diferenciação de significados, não uma perda de significados. Se o esquecimento for total, como se o indivíduo nunca tivesse aprendido um certo conteúdo, é provável que a aprendizagem tenha sido mecânica e não significativa.

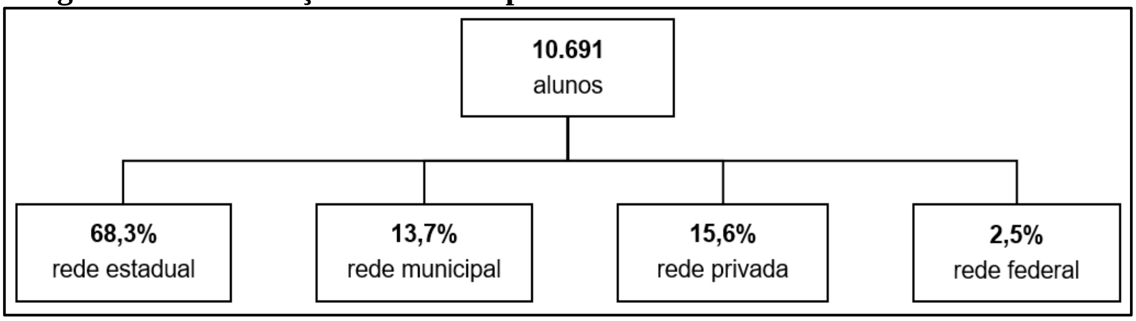
Um indicador que mostra o quanto se torna urgente tal apropriação da aprendizagem significativa pelos educandos brasileiros são os dados divulgados pelo Ministério da Educação[1] referentes ao Programa Internacional de Avaliação de Estudantes (PISA)[2]. O PISA é um exame amostral aplicado a cada três anos pela Organização para Cooperação e Desenvolvimento Econômico (OCDE)[3], cuja finalidade é mensurar até que ponto os jovens na faixa etária dos 15

[1] http://portal.mec.gov.br. Acesso em: 05 de dezembro de 2019.
[2] https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa/resultados. Acesso em: 05 de dezembro de 2019.
[3] https://www.oecd.org/pisa/data/. Acesso em: 05 de dezembro de 2019.

anos, próximos ao final da educação básica, adquiriram conhecimentos e habilidades fundamentais para exercer sua participação na vida social e econômica.

Até o presente momento foram aplicados sete exames do PISA, datados entre 2000 a 2018, que contaram com a participação de aproximadamente 600 mil estudantes dentre 79 países. Nesta última edição do exame, o Brasil participou com 10.691 alunos de escolas públicas e privadas, nascidos em 2002, selecionados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Como apresentado na figura 1, quase 70% destes 10.691 participantes pertenciam às redes estatuais de ensino.

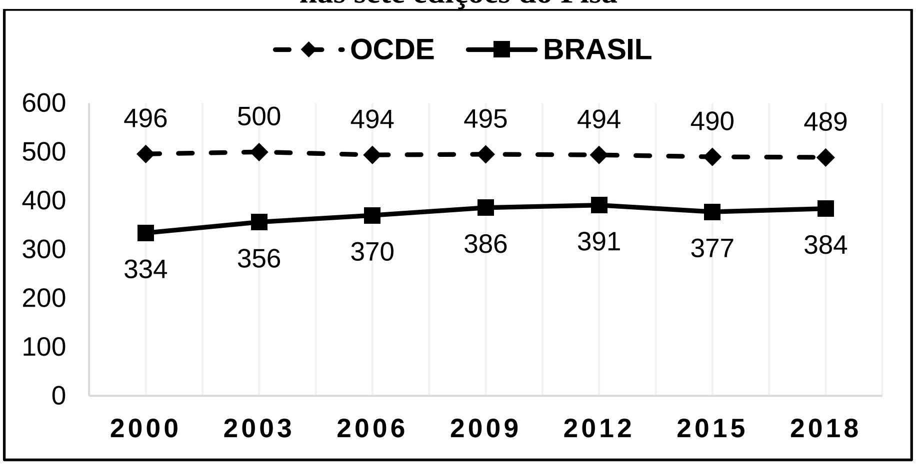
Figura 1 – Distribuição dos alunos por rede educacional no exame do Pisa 2018



Fonte: Adaptado de Ministério da Educação – Brasil no Pisa 2018
Acesso em: 05/12/2019

Ao analisar os resultados na prova de matemática com os dados apresentados pela OCDE nos sete exames aplicados, percebe-se que os alunos brasileiros têm desempenho abaixo da média das outras 78 nações participantes como observa-se no Gráfico 1.

Gráfico 1 – Desempenho médio dos educandos brasileiros em matemática nas sete edições do Pisa



Fonte: Adaptado https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa/resultados
Acesso em: 05/12/2019

A oportunidade de participar do programa do PISA permite ao Brasil comparar seu resultado com o de outros países mais desenvolvidos e, assim, ser inspirado à adoção de medidas que mudem a atual qualidade de ensino disposta aos jovens educandos brasileiros, promovendo a mudança da realidade hodierna. Ressalta-se que a qualidade da educação está preconizada pelo investimento na formação dos profissionais docentes. Estes, quando bem preparados e motivados, são parte integrante do alicerce que garante uma reconstrução de modelo educacional.

Ainda no Gráfico 1, percebe-se pelo período de 19 anos, contados a partir do ano 2000, que os educandos brasileiros tiveram uma evolução acerca do domínio da matemática e das ações que exprimem como os indivíduos abordam e resolvem um problema que envolva a matéria. Por conseguinte, relatórios da OCDE (2019) analisam que é imprescindível que se continue investindo em políticas públicas para a melhoria da educação, de forma a garantir as aprendizagens básicas onde o aluno ganhe proficiência e desenvolva a capacidade de interpretar, formular e empregar os conhecimentos matemáticos na resolução de problemas do cotidiano. Agora, como garantir as aprendizagens básicas e aumentar a proficiência dos educandos?

Com o objetivo geral de motivar o estudante na busca pela compreensão da Matemática, é importante desenvolver estratégias para o ensino desta ciência de forma contextualizada. Particularmente, no que diz respeito ao ensino da Estatística e da Probabilidade, a busca pela aprendizagem significativa torna-se premente, uma vez que vivemos em uma realidade cada vez mais tecnológica, interconectada e geradora de informações.

Segundo Lopes (2008, p. 59) “o desenvolvimento da estatística e da probabilidade, nas escolas básicas, tem sido alvo de pesquisas em algumas partes do mundo, e muitos pesquisadores publicam trabalhos a respeito, procurando justificar a relevância do assunto”. A necessidade da efetiva apropriação deste conhecimento, desde o Ensino Fundamental, se impõe como um dos quesitos para a formação do cidadão das atuais e futuras gerações. Estes, deverão ser cada vez mais capazes de interpretar e analisar as informações com o intuito de tomar decisões com o menor grau de incerteza.

Nas atividades do magistério, observa-se que a abordagem empregada no ensino da estatística descritiva e da probabilidade, na maioria dos livros didáticos utilizados na Educação Básica, se valem frequentemente de situações hipotéticas e desarticuladas da realidade do aluno, o que contribui pouco com o processo ensino-aprendizagem do tema. Vale ressaltar que não se pretende discordar da didática e das atividades fictícias presentes nos livros, uma vez que as problematizações hipotéticas utilizadas nesses, cumprem com o objetivo de transmitir o conhecimento. Todavia, os exercícios nem sempre despertam o interesse do aluno e, portanto, não deveriam ser o principal ou único recurso de explanação e consolidação da aprendizagem.

O ritmo das transformações na sociedade moderna mostra a necessidade de repensar o ensino da matemática. Os profissionais da educação que se dedicam ao ensino das ferramentas e conceitos matemáticos, devem reformular suas práticas docentes e exercitar atividades, por meio de uma sequência didática que trabalhe a investigação, a reflexão e a análise crítica para a resolução de problemas atrelados às demandas cotidianas. Segundo Cabral (2017), a sequência didática é um conjunto de atividades estruturadas de maneira sistemática pelo professor que propiciam um ambiente onde o processo ensino-aprendizagem se dá pela contextualização, criação, adaptação, avaliação e elaboração de métodos, cuja última finalidade é minimizar as dificuldades de aprendizagem.

Assim, com o objetivo específico de oportunizar que esta aprendizagem seja significativa, propõem-se o desenvolvimento de um projeto ou trabalho investigativo, em que o educando fará a modelagem de um conjunto de dados por tabelas, quadros e gráficos para, posteriormente, fazer análises e construir inferências que podem ser reforçadas ou credibilizadas por cálculos probabilísticos, caso seja conveniente. Ressalta-se que esse trabalho deve ser implementado com dados oriundos de uma pesquisa factível e que desperte o interesse do aluno.

Como exemplo de projeto que cumpra este fim, apresenta-se nesta dissertação uma proposta de atividade aberta, com algumas possibilidades de encaminhamento, onde são usados dados retirados do site do Ministério da Saúde sobre o número de indivíduos infectados pelo vírus da Dengue no Brasil de 2009 a 2019. O foco desta atividade, que se apoia na metodologia da sequência didática, é a aplicação dos conhecimentos adquiridos sobre estatística descritiva e a teoria da probabilidade. Espera-se também que o educando proponha soluções e argumentações consistentes às indagações que possam surgir pelas informações expostas durante a modelagem dos dados. Neste ponto, o professor passa a ser um mero consultor para tirar dúvidas acerca dos conceitos matemáticos envolvidos. De acordo com Camargo e Daros (2018, p. 23)

> Sabe-se que, em espaços nos quais os professores assumem a centralidade do processo e se apresentam como detentores de todo o conhecimento, acaba-se por impossibilitar a participação mais ativa dos estudantes e, ainda, se instaura o medo de errar, de arriscar e de participar.

A aprendizagem efetiva ocorre quando o estudante é capaz de aplicar o conhecimento transmitido de modo a evidenciar o alcance de sua autonomia. Nesse sentido, propostas de atividades baseadas em modelagem matemática vêm ao encontro do objetivo de tornar o estudante um agente ativo no processo ensino-aprendizagem. Por último, as atividades geradas durante uma modelagem, interdisciplinarmente ou transdisciplinarmente, tem um forte potencial a ser desenvolvido.

A demanda por ações pedagógicas e didáticas que contribuam no contexto da Educação Matemática, mostram a necessidade de programas de pós-graduação, como o PROFMAT, e de pesquisas afins. Por um fichamento realizado previamente com dissertações presentes no repositório deste programa e em sites de revistas científicas, constata-se que os temas "Estatística" e "Modelagem" são abordados em diversas propostas metodológicas de ensino.

Dentre os trabalhos pesquisados no repositório do PROFMAT, elencamos alguns que conversam com a proposta apresentada nesta dissertação:

- Pinho (2013), constrói um elo entre a realidade dos alunos e a construção do conhecimento matemático, abordando conceitos da estatística e da probabilidade e utilizando uma proposta de atividade em que a turma escolhe um tema para estruturar uma pesquisa dentro de sua comunidade escolar. Em seguida, formulam um questionário a ser aplicado para a coleta dos dados que serão modelados em tabelas e gráficos. O último passo é a elaboração de um resumo com o objetivo da pesquisa, inferências construídas e conclusões. Ao final da atividade, a construção do conhecimento terá ocorrido pelas próprias experiências e esse resultado é exposto na aplicação dos conhecimentos estatísticos durante a construção do texto;
- Duarte (2013), coloca a importância das metodologias ativas para o ensino da matemática, e ressalta que foram desenvolvidas para se adaptarem à mudança de comportamento dos estudantes. Em uma turma de 3ª série do ensino médio, é abordado o ensino da estatística descritiva baseada no cotidiano dos alunos, suas vivências tanto na escola como no ambiente que vivem e convivem. Sua conclusão é que este tipo de proposta tem um resultado satisfatório causando interesse nos educandos e o consequente aprendizado;
- Rios (2014), motiva o professor de matemática da educação básica a lecionar os conteúdos de estatística e probabilidade de forma a dar mais sentido às aprendizagens. A ideia, é trabalhar com experimentos mais concretos que levem

o aluno a aumentar sua capacidade de entendimento sobre aleatoriedade, população, amostra, análise e representatividade das conclusões tiradas do projeto de pesquisa. Em suma, é possível firmar que muito mais pode ser ensinado na Educação Básica, inclusive pontos da inferência estatística que quase nunca são trabalhados nesta fase do ensino. Outrossim, as propostas de ensino devem ser executadas de maneira lúdica, motivacional e significativa para o aluno;

- Silva (2015), reflete sobre as práticas pedagógicas do professor de matemática e as contribuições da aprendizagem significativa no ensino de estatística utilizando o *software* livre Calc. Sua pesquisa evidenciou caminhos metodológicos a serem desenvolvidos por professores de matemática na educação básica que contribuam na interpretação de dados do cotidiano através de tabelas, gráficos etc.;
- Souza (2018), verificou o aprendizado dos conteúdos de estatística e probabilidade por parte de alguns alunos na 1ª, 2ª e 3ª séries do ensino médio. Em sua proposta, aplicou uma avaliação diagnóstica com 12 questões que testariam o conhecimento. A autora coloca que o ideal para se ensinar estatística, seria uma construção contínua e gradativa que aprofundasse o assunto em cada uma das três etapas do ensino médio. As aulas devem ser diferenciadas com a aplicação do conhecimento no cotidiano por meio de oficinas que trabalhem com material concreto (dados, cartas de baralho) e recursos tecnológicos.

Dois artigos, publicados na Revista Eletrônica Vidya, também chamam atenção sobre o assunto:

- Lopes et al. (2013), apresentam e promovem reflexões sobre três atividades e projetos que tratam do ensino da estatística e da probabilidade, desenvolvidos em dissertações de mestrado nos três níveis da Educação Básica (Educação Infantil, Ensino Fundamental e Ensino Médio). As análises são construídas de acordo com o potencial das abordagens de ensino e pelos processos cognitivos empregados pelos alunos ao aprenderem Estatística pela investigação. Observou-se também o desenvolvimento dos alunos durante a aplicação de atividades que fizessem a modelagem de dados pela Estatística. Os resultados mostram que motivar os estudantes de todos os níveis de ensino, por vivências em métodos de coleta e análise de dados, facilita o desenvolvimento de um raciocínio focado na visão estatística/probabilística. Percebe-se que os alunos aprendem a investigar e a tirar conclusões próprias. Além disso, conclui-se que nas práticas docentes, é necessário desenvolver atividades e projetos pautados pela realidade do educando como forma de abordar os conteúdos da Estatística. O objetivo é possibilitar que aprendam a fazer uma leitura crítica de sua realidade;

- Burak e Penteado (2019), apresentam os resultados de uma pesquisa qualitativa com 31 trabalhos constantes, nos Anais do Encontro Paranaense de Modelagem na Educação Matemática (EPMEM), sobre as práticas no ensino de matemática pela modelagem com o objetivo de divulgar os resultados dos projetos desenvolvidos. Pela análise desses trabalhos, infere-se que as atividades e projetos aplicados em sala de aula, pela metodologia da modelagem matemática, resultam em práticas que auxiliam os alunos na aprendizagem significativa dos conteúdos matemáticos abordados.

A partir desse estudo, entende-se que a investigação, a exploração de dados, a problematização da realidade, a modelagem e o uso de tecnologias são relevantes no processo do ensino e contribuem para a construção e a fixação do conhecimento. Sob esse contexto, as atividades desenvolvidas com base nesses cinco pilares ajudam a suprir uma demanda de ações que contribuem para a aprendizagem significativa da Matemática. Nessa conjuntura, identifica-se que o que diferencia esse trabalho dos demais acima citados é a natureza da modelagem que se propõem, construída com dados provenientes de sites de instituições governamentais confiáveis, para então proceder às análises e produzir informações e conhecimentos que dão foco científico ao trabalho desenvolvido pelos educandos. Dessa forma, busca-se a valoração das competências cognitivas que não se evidenciam na forma convencional do ensino e que deixam de trabalhar a proatividade, a colaboração e a interação pessoal.

Nesta dissertação, coloca-se o emprego da modelagem não como fim, mas como meio para o processo ensino-aprendizagem da Estatística e Probabilidade. Soma-se a esse propósito a possibilidade de que as atividades desenvolvidas podem dar ao educando a percepção de como o saber matemático está presente nas diversas áreas do conhecimento.

Incentivar e dar sentido ao aprendizado, com propostas de modelagem que façam uso de diversas ferramentas matemáticas e o emprego das tecnologias que estão à disposição do ensino, implica em uma ruptura com a linearidade da atual prática educativa. Usar do cotidiano, dos fatos que permeiam o mundo globalizado, engajado em constante inovação tecnológica, possibilitam desenvolver uma prática pedagógica por meio de experimentações concretas. Isto é especialmente factível no que diz respeito à análise de dados auxiliada por recursos computacionais.

Segundo D'Ambrósio (2009, p. 86) "A responsabilidade maior do professor vai, portanto, além da sua disciplina específica. Mas, hoje cidadania implica conhecimento". Nessa perspectiva, é preciso que na Educação Básica se desenvolva um currículo que propicie ao aluno a aquisição do conhecimento e de habilidades para que este exerça sua cidadania em um mundo tecnologicamente mais integrado e bombardeado por um grande volume de informação gerada e disseminada pelas mais diversas fontes midiáticas. Como colocado por D'Ambrósio (2009, p. 87) "A educação para cidadania, que é um dos grandes objetivos da educação de hoje, exige uma apreciação do conhecimento moderno, impregnado de ciência e tecnologia".

2 A BNCC E O ENSINO DA ESTATÍSTICA E DA PROBABILIDADE

Na década de 80, o mundo enfrentava um quadro de enormes diferenças econômicas entre as nações, guerras, lutas civis e degradação do meio-ambiente na busca do desenvolvimento econômico. Neste cenário, em 1990 foi realizada a Conferência Mundial sobre Educação para Todos na cidade de Jomtien, Tailândia. Nesta conferência, começou a ser discutido um novo entendimento sobre Educação para satisfazer as necessidades básicas de aprendizagem que levasse a sociedade mundial a superar seus problemas.

Seguindo este movimento, o governo brasileiro promulgou a Lei nº 9.394 em 20 de dezembro de 1996, que estabelece as diretrizes e bases da Educação Básica nacional. A partir desta lei, o Conselho Nacional de Educação (CNE) trabalhou 21 anos e elaborou a Base Nacional Comum Curricular (BNCC), que foi promulgada em dezembro de 2017 através da resolução Conselho Nacional de Educação/Conselho Pleno (CNE/CP) nº 2. Este documento de caráter normativo, trata dos conhecimentos essenciais a serem trabalhados na Educação Básica brasileira nos níveis da Educação Infantil e Ensino Fundamental. Posteriormente em dezembro de 2018, houve uma nova resolução CNE/CP nº 4, que instituiu a Base Nacional Comum Curricular na etapa do Ensino Médio (BNCC-EM). O objetivo é garantir o direito pleno à aprendizagem e promover, de forma plural, a igualdade no sistema educacional conforme prevê o Plano Nacional de Educação (PNE).

Ao longo da Educação Básica, espera-se que o estudante desenvolva 10 competências gerais. Segundo o Ministério da Educação pela BNCC (2018, p. 08) “competência é definida como a mobilização de conhecimentos (conceitos e procedimentos), habilidades (práticas, cognitivas e socioemocionais), atitudes e valores para resolver demandas complexas da vida cotidiana, do pleno exercício da cidadania e do mundo do trabalho”. De acordo com Ministério da Educação pela BNCC (2018, p. 09 e 10), as 10 competência são:

> 1. Valorizar e utilizar os conhecimentos historicamente construídos sobre o mundo físico, social, cultural e digital para entender e explicar a realidade, continuar aprendendo e colaborar para a construção de uma sociedade justa, democrática e inclusiva.
> 2. Exercitar a curiosidade intelectual e recorrer à abordagem própria das ciências, incluindo a investigação, a reflexão, a análise crítica, a imaginação e a criatividade, para investigar causas, elaborar e testar hipóteses, formular e

> resolver problemas e criar soluções (inclusive tecnológicas) com base nos conhecimentos das diferentes áreas.
> 3. Valorizar e fruir as diversas manifestações artísticas e culturais, das locais às mundiais, e também participar de práticas diversificadas da produção artístico-cultural.
> 4. Utilizar diferentes linguagens – verbal (oral ou visual-motora, como Libras, e escrita), corporal, visual, sonora e digital –, bem como conhecimentos das linguagens artística, matemática e científica, para se expressar e partilhar informações, experiências, ideias e sentimentos em diferentes contextos e produzir sentidos que levem ao entendimento mútuo.
> 5. Compreender, utilizar e criar tecnologias digitais de informação e comunicação de forma crítica, significativa, reflexiva e ética nas diversas práticas sociais (incluindo as escolares) para se comunicar, acessar e disseminar informações, produzir conhecimentos, resolver problemas e exercer protagonismo e autoria na vida pessoal e coletiva.
> 6. Valorizar a diversidade de saberes e vivências culturais e apropriar-se de conhecimentos e experiências que lhe possibilitem entender as relações próprias do mundo do trabalho e fazer escolhas alinhadas ao exercício da cidadania e ao seu projeto de vida, com liberdade, autonomia, consciência crítica e responsabilidade.
> 7. Argumentar com base em fatos, dados e informações confiáveis, para formular, negociar e defender ideias, pontos de vista e decisões comuns que respeitem e promovam os direitos humanos, a consciência socioambiental e o consumo responsável em âmbito local, regional e global, com posicionamento ético em relação ao cuidado de si mesmo, dos outros e do planeta.
> 8. Conhecer-se, apreciar-se e cuidar de sua saúde física e emocional, compreendendo-se na diversidade humana e reconhecendo suas emoções e as dos outros, com autocrítica e capacidade para lidar com elas.
> 9. Exercitar a empatia, o diálogo, a resolução de conflitos e a cooperação, fazendo-se respeitar e promovendo o respeito ao outro e aos direitos humanos, com acolhimento e valorização da diversidade de indivíduos e de grupos sociais, seus saberes, identidades, culturas e potencialidades, sem preconceitos de qualquer natureza.
> 10. Agir pessoal e coletivamente com autonomia, responsabilidade, flexibilidade, resiliência e determinação, tomando decisões com base em princípios éticos, democráticos, inclusivos, sustentáveis e solidários.

Como esse trabalho se debruça sobre o ensino da estatística e da probabilidade, a ênfase será nas competências específicas de Matemática e suas Tecnologias no Ensino Médio, pois demandam um aprendizado de matemática mais elaborado, que exige um conhecimento teórico e raciocínio criativo para a solução de problemas reais. Como coloca o Ministério da Educação pela BNCC (2018, p. 470)

> [...] os estudantes devem utilizar conceitos, procedimentos e estratégias não apenas para resolver problemas, mas também para formulá-los, descrever dados, selecionar modelos matemáticos e desenvolver o pensamento computacional, por meio da utilização de diferentes recursos da área.

No Ensino Médio, a BNCC prevê a consolidação e o aprofundamento dos aprendizados essenciais. Trata-se de aprofundar o pensamento computacional, iniciado no Ensino Fundamental, que envolve habilidades para raciocinar e resolver problemas aplicáveis em sua realidade, tendo o computador como ferramenta auxiliar pelo uso de *softwares* e de ampliar o letramento matemático, que segundo o Ministério da Educação pela BNCC (2018), trata das competências e habilidades que devem ser trabalhadas com o educando, possibilitando a este, que seja capaz de utilizar os conceitos e as ferramentas matemáticas para raciocinar, representar, comunicar e argumentar matematicamente estabelecendo conjecturas em vários contextos na perspectiva de privilegiar o esforço produtivo. Neste ponto, é necessário estimular processos mais complexos de reflexão e abstração que permitam aos estudantes formular e resolver problemas em diversos contextos com mais autonomia e recursos matemáticos. Para isso, os estudantes devem desenvolver habilidades relacionadas aos processos de investigação, de construção de modelos e de resolução de problemas.

Na matemática, cinco competências serão exploradas ao longo do Ensino Médio. Para este trabalho, será ressaltada a competência número três que segundo o Ministério da Educação pela BNCC (2018, p. 27)

> [...] utiliza estratégias, conceitos e procedimentos matemáticos, em seus campos da [...] Probabilidade e Estatística para interpretar, construir modelos e resolver problemas em diversos contextos, analisando a plausibilidade dos resultados e a adequação das soluções propostas, de modo a construir argumentação consistente.

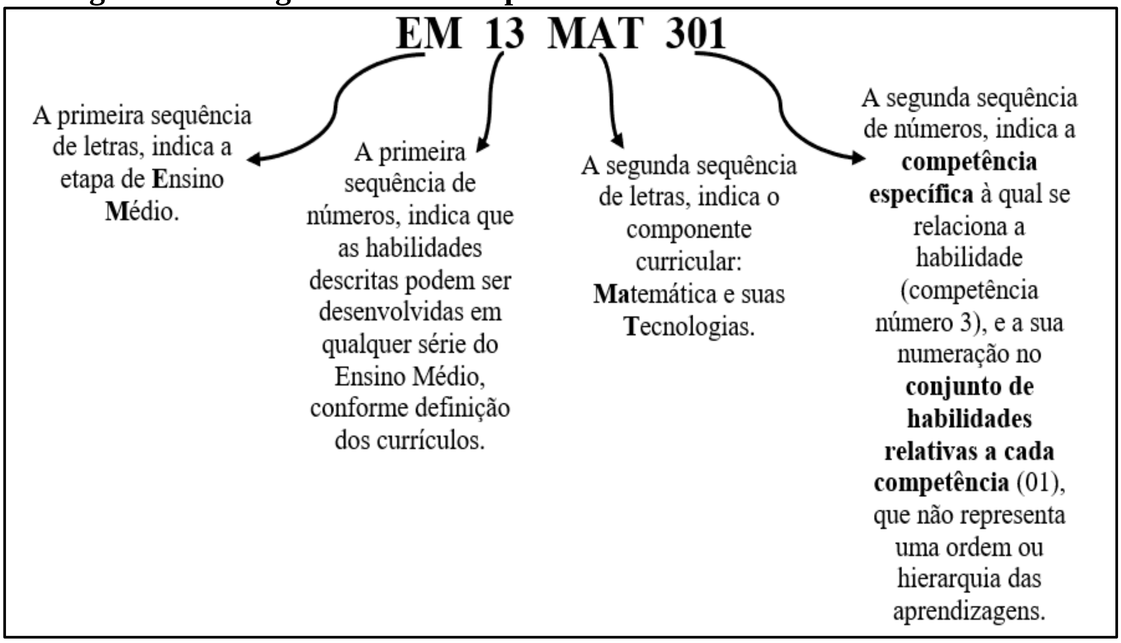
Além das competências, a BNCC estabeleceu várias habilidades que precisam ser estimuladas nos discentes. São habilidades gerais, que devem ser o objetivo do ensino da matemática como um todo. Porém, destacam-se algumas habilidades específicas da competência número três que são, segundo o Ministério da Educação pela BNCC (2018, p. 28 e 29)

> **(EM13MAT301)** Resolver e elaborar problemas do cotidiano, da Matemática e de outras áreas do conhecimento, que envolvem equações lineares simultâneas, usando técnicas algébricas e gráficas, incluindo ou não tecnologias digitais [...]
> **(EM13MAT303)** Resolver e elaborar problemas envolvendo porcentagens em diversos contextos [...]
> **(EM13MAT304)** Resolver e elaborar problemas com funções exponenciais nos quais é necessário compreender e interpretar a variação das grandezas envolvidas [...]
> **(EM13MAT305)** Resolver e elaborar problemas com funções logarítmicas nos quais é necessário compreender e interpretar a variação das grandezas envolvidas [...]

> **(EM13MAT310)** Resolver e elaborar problemas de contagem envolvendo diferentes tipos de agrupamento de elementos, por meio dos princípios multiplicativo e aditivo, recorrendo a estratégias diversas como o diagrama de árvore. [...]
> **(EM13MAT311)** Resolver e elaborar problemas que envolvem o cálculo da probabilidade de eventos aleatórios, identificando e descrevendo o espaço amostral e realizando contagem das possibilidades.
> **(EM13MAT312)** Resolver e elaborar problemas que envolvem o cálculo de probabilidade de eventos em experimentos aleatórios sucessivos.
> **(EM13MAT313)** Resolver e elaborar problemas que envolvem medições em que se discuta o emprego de algarismos significativos e algarismos duvidosos, utilizando, quando necessário, a notação científica. [...]
> **(EM13MAT316)** Resolver e elaborar problemas, em diferentes contextos, que envolvem cálculo e interpretação das medidas de tendência central (média, moda, mediana) e das de dispersão (amplitude, variância e desvio padrão).

Estas habilidades citadas podem ser desenvolvidas em qualquer etapa do Ensino Fundamental II e do Ensino Médio, e buscam representar as aprendizagens essenciais que devem ser garantidas aos estudantes. São identificadas por um código alfanumérico cuja formação é subdividida em quatro partes bem definidas como observa-se no esquema abaixo.

Figura 2 – Código alfanumérico para as habilidades constantes na BNCC



Fonte: Adaptado BNCC (2018, p.34)

3 METODOLOGIA

3.1 Modelagem Matemática

A modelagem matemática estuda um sistema para descrever ou prever o comportamento dos fenômenos naturais. Exemplos clássicos de modelagens matemáticas que transformaram a ciência moderna são as leis de Isaac Newton, com as quais o homem aumentou significativamente o seu poder de interferir na realidade. Mais do que isso, no entanto, a modelagem pode ser vista como um dos caminhos que oportuniza ao educando aplicar os conceitos matemáticos adquiridos. Desta forma, o aprendizado se dará por métodos mais atrativos e menos repetitivos.

Como coloca Bassanezi (2002, p. 17)

> No caso específico da Matemática, é necessário buscar estratégias alternativas de ensino-aprendizagem que facilitem sua compreensão e utilização. A modelagem matemática, em seus vários aspectos, é um processo que alia teoria e prática, motiva seu usuário na procura do entendimento da realidade que o cerca e na busca de meios para agir sobre ela e transformá-la. Nesse sentido, é também um método científico que ajuda a preparar o indivíduo para assumir seu papel de cidadão. A educação inspirada nos princípios da liberdade e da solidariedade humana tem por fim o preparo do indivíduo e da sociedade para o domínio dos recursos científicos e tecnológicos que lhes permitem utilizar as possibilidades e vencer as dificuldades do meio.

Na forma clássica do ensino de matemática, o aluno aprende a resolver problemas, mas costuma esquecer o que aprendeu com o tempo. A modelagem ensina o aluno a solucionar problemas matemáticos pré-elaborados, bem como a resolver problemas da vida com o uso da matemática, o que aumenta o poder de retenção dos saberes e o alcance do aprendizado. É necessário criar condições para ter uma participação mais ativa dos alunos. A inovação abre caminhos que possibilitam estabelecer uma aprendizagem significativa entre os diversos saberes de forma progressiva e estimulante. O desenvolvimento de estratégias que garantam a organização de um aprendizado mais interativo, passa pela “compreensão das chamadas Metodologias Ativas de Aprendizagem, que nada mais são do que métodos para tornar o estudante protagonista do seu processo de aprendizagem, e não mais elemento passivo na recepção de informações” (CAMARGO e DAROS, 2018, p. 3).

Neste sentido, a modelagem passa a ser pensada em termos da metodologia da sequência didática que segundo Zabala (1998, p. 18) é “um conjunto de atividades ordenadas, estruturadas e articuladas para a realização de certos objetivos educacionais, que têm um princípio e um fim conhecidos tanto pelos professores como pelos alunos”. Cabral (2017) coloca que estas atividades organizadas e gerenciadas pelo professor, têm por objetivo, promover a aprendizagem como elos conectados de uma corrente que permitem as articulações entre si.

Segundo Barbosa (2001, p. 3) a “modelagem pode ser definida em termos dos propósitos e interesses subjacentes à sua implementação, conduzindo a implicações conceituais e curriculares”. Na educação básica, a modelagem pode ser usada como uma forma instigante e desafiadora de ensinar matemática pela pesquisa, formando uma consciência criativa, científica e capaz de solucionar problemas cada vez mais complexos. Segundo Bassanezi (2002, p. 32 e p. 33)

> Uma série de pontos podem ser levantados para destacar a relevância da modelagem matemática quando utilizada como instrumento de pesquisa:
> • Pode estimular novas ideias e técnicas experimentais;
> • Pode dar informações em diferentes aspectos dos inicialmente previstos;
> • Pode ser um método para se fazer interpolações, extrapolações e previsões;
> • Pode sugerir prioridades de aplicações de recursos e pesquisas e eventuais tomadas de decisão;
> • Pode preencher lacunas onde existem falta de dados experimentais;
> • Pode servir como recurso para melhor entendimento da realidade;
> • Pode servir de linguagem universal para compreensão e entrosamento entre pesquisadores em diversas áreas do conhecimento.
> A modelagem matemática, com toda sua abrangência e poder de síntese, é por excelência o método científico usado nas ciências factuais [...]

A modelagem é “uma oportunidade para os alunos indagarem situações por meio da matemática sem procedimentos fixados previamente e com possibilidades diversas de encaminhamento” (BARBOSA, 2001, p. 5). Ela é capaz de estimular, concomitantemente, a busca por distintos caminhos para a solução de um problema e o aprendizado duradouro e prazeroso da matemática.

3.2 Sequência Didática

O ato de ensinar, é desafiador por natureza. Especificamente no que diz respeito ao ensino de Matemática, este ato torna-se ainda mais desafiador dada a ausência de uma boa base de conhecimento em matemática por parte dos discentes. Principalmente por se tratar de uma matéria cujos ensinamentos são acumulativos. Ou seja, se durante um ano letivo o educando não abstraiu significativamente os conteúdos ministrados de Matemática, é possível que tenha dificuldade para aprender os conteúdos que serão propostos posteriormente. Este é um dos motivos pelo qual muitos criam bloqueios que atrapalha a depreensão do aprendizado matemático.

Neste ponto, modelos metodológicos alternativos procuram minimizar as dificuldades de se ensinar. “As preocupações desses modelos metodológicos estão dirigidas às formas de ensinar que buscam explicitar a inteligibilidade do objeto ensinado na percepção do aluno” (CABRAL, 2017, p. 9). O educador precisa se fazer entender diversificando sua didática e fugindo da tradicional tríade: definição, exemplo e exercício.

Inicialmente, a metodologia da sequência didática foi desenvolvida para trabalhar o aprendizado de gênero textual oral ou escrito. Sua organização é construída em quatro fases bem definidas que são: a apresentação da situação de ensino, a produção inicial, os módulos e a produção final. Segundo Cabral (2017, p. 33 e 34)

> Na primeira fase, os alunos recebem do professor uma descrição minuciosa da relevância do projeto de ensino em questão bem como dos objetivos, estrutura e condições coletivas de produção dos saberes envolvidos. Já a segunda fase, qual seja, a produção inicial, guarda as intervenções que visam diagnosticar as capacidades já adquiridas pelos alunos em relação ao gênero objeto de ensino e, além disso, procura adequar às ações de ensino posteriores a partir das quais se pretende atingir os objetivos de aprendizagem. Após essa fase diagnóstica dos sujeitos, vem a terceira fase – desenvolvimento dos módulos – na qual serão ministradas as oficinas que se constituem em diversas atividades, relativas ao desenvolvimento das capacidades de linguagem, envolvendo as três práticas linguísticas: leitura, produção e análise da língua. O número de módulos/oficinas é flutuante e deve se adequar ao suprimento das dificuldades encontradas pelos alunos na escrita inicial do gênero objeto de estudo. [...] Após os módulos, segue-se a quarta fase - a produção final, na qual o aluno coloca em prática os conhecimentos adquiridos e, juntamente com o professor, avaliam os progressos alcançados.

A estrutura metodológica da sequência didática em si, no entanto, é perfeitamente adaptável para o ensino da Matemática através de uma atividade investigativa onde os educandos colocam seus conhecimentos em prática. No capítulo 6 apresenta-se uma proposta de atividade construída nos moldes desta metodologia, acima citada, e adaptada para o ensino da Matemática.

Ao elaborar uma atividade neste princípio metodológico, o professor deve valorizar o conhecimento prévio dos alunos, estipular níveis de complexidades, utilizar de temas diversificados e centrados na problematização, gerando assim, um ambiente que favoreça a interação e a sistematização dos saberes.

Tornar a sala de aula um espaço mais democrático, tirando do professor a responsabilidade de ser a principal fonte de transmissão do conhecimento, leva a uma perspectiva de romper com o tradicional, de estimular a reflexão teórica e abrir espaço para as vivências e experiências ganharem protagonismo. O objetivo é gerar uma autonomia pedagógica e uma agitação intelectual, que propicie a aprendizagem pela prática sempre embasada pelas diversas Ciências, mantendo a cooperação entre os educandos envolvidos e focando nas aprendizagens matemáticas a cada etapa da sequência didática.

4 FUNDAMENTAÇÃO TEÓRICA

4.1 Estatística Descritiva

A estatística descritiva é "o ramo da estatística que envolve a organização, o resumo e a representação dos dados" (FARBER e LARSON, 2015, p. 5). Os profissionais das mais diversas áreas, fazem uso dessa ferramenta para modelar dados coletados durante uma pesquisa através da descrição de suas características e comportamentos. Sua importância é indiscutível no campo das Ciências Exatas, das Ciências da Natureza e das Ciências Humanas. Prova disso, é só observar que ela é uma matéria presente na grade curricular dos mais variados cursos de graduação como: medicina, psicologia, economia, publicidade, administração, física, química, engenharias de uma forma geral entre outros.

Durante seu trabalho, o pesquisador necessitará organizar os dados coletados transformando-os em informações claras e precisas. O intuito é possibilitar uma análise e comparação com outras informações já conhecidas a partir das quais se podem tomar decisões. Conforme define Andrade et al. (2015, p. 23)

> Para que se obtenham resultados válidos, o investigador deve seguir todos os passos que definem o método estatístico de resolução de problemas:
> 1. Identificar corretamente o problema em análise. Mesmo em estudos exploratórios cujo objetivo é identificar possíveis relações entre as características dos indivíduos sem que, a princípio, se defina um modelo regulador dessas relações, é necessário identificar o problema para o qual se pretendem encontrar respostas.
> 2. Recolher a informação necessária, relevante para o problema em estudo, em tempo útil e tão completa quanto possível. Esta informação poderá consistir em dados primários, recolhidos através de um questionário, ou dados secundários, recolhidos e publicados através de outra fonte de informação.
> 3. Classificar e organizar os dados, por exemplo, através da codificação e criação de uma base de dados em suporte informático. Uma vez ultrapassada esta fase, é possível reduzir a quantidade de informação, filtrando pormenores irrelevantes através de medidas de estatística descritiva com tabelas, técnicas gráficas, medidas de tendência central, dispersão e concentração.
> 4. Análise dos dados e apresentação dos resultados: identificar relações, testar hipóteses, definir modelos com a ajuda de métodos estatísticos apropriados.
> 5. Tomar a decisão mais adequada, ponderando as possíveis opções face aos objetivos inicialmente propostos. A qualidade da informação recolhida e as capacidades do investigador determinam, em grande parte, a adequabilidade das opções propostas.

Uma boa análise exploratória de dados, deve ser feita através de tabelas, técnicas gráficas, medidas de posição (médias, moda, mediana e separatrizes) e variabilidade (variância, desvio médio e desvio-padrão). Aqui, o principal objetivo é entender o que estes dados nos dizem. Segundo Farias (2020, p. 1) "Estatística é a ciência da aprendizagem a partir dos dados".

Neste capítulo, apresentaremos notações, definições, tabelas e técnicas gráficas que organizam um conjunto de dados e nos levam a compreender as características de uma população.

4.1.1 População

A população é o conjunto formado por elementos que possuem uma característica comum, a qual se tem o interesse de estudar.

São exemplos de uma população:

- as pessoas que foram diagnosticadas com câncer no Brasil nos últimos 20 anos;
- os 22 atletas que estão em campo ao término de um jogo de futebol;
- os carros produzidos por uma montadora no Brasil nos últimos 15 anos.

4.1.2 Amostra

A amostra é um subconjunto formado por elementos selecionados aleatoriamente de uma população.

São exemplos de uma amostra:

- 1.500 pessoas selecionadas aleatoriamente, dentre todas as que foram diagnosticadas com câncer no Brasil nos últimos 20 anos;
- 4 atletas sorteados aleatoriamente para fazer o exame antidoping, dentre os 22 que estão em campo ao término de um jogo de futebol.

4.1.3 Variáveis qualitativas

As variáveis qualitativas descrevem as características dos elementos de uma população. Vejamos suas classificações com alguns exemplos:

- ordinal → classe social (A, B, C, D), estágio da doença (inicial, intermediário, terminal), entre outros;
- nominal → sexo (masculino, feminino), raça (branco, pardo, negro, amarelo, indígena), entre outros.

4.1.4 Variáveis quantitativas

As variáveis quantitativas mensuram as características dos elementos de uma população através de contagem ou medida. São classificadas em discretas, quando a mensuração tem uma quantidade finita de resultados que podem ser enumeráveis, e em contínuas, quando a mensuração tem uma quantidade finita de resultados cujos valores pertencem a intervalos na reta real.

A seguir, observa-se alguns exemplos:

- discreta → quantidade de pessoas infectadas por um vírus, quantidade de filhos de um casal, quantidade de livros em uma biblioteca, entre outros;
- contínua → massa corporal de um indivíduo, altura de uma árvore, unidade monetária, tempo de uma volta de um carro de fórmula 1, volume em uma garrafa, entre outros.

4.1.5 Dados

Os dados são as informações coletadas das características de uma população por meio das observações, podendo ser numéricas ou não.

São exemplos:

- das 22.169 pessoas diagnosticadas com a doença covid-19 no território brasileiro até a data de 12 de abril de 2020, 1.223 entraram em óbito (dado quantitativo)[4];
- segundo o censo de 2015, a cor ou raça da população brasileira com base na autodeclaração são branca, parda, preta, amarela e indígena (dado qualitativo)[5].

[4] Fonte: https://covid.saude.gov.br/ Acesso em: 13 de abril de 2020.

[5] Fonte: https://educa.ibge.gov.br/jovens/conheca-o-brasil/populacao/18319-cor-ou-raca.html Acesso em: 28 de março de 2020.

4.1.6 Rol

Rol é a organização dos dados de uma amostra em ordem crescente ou decrescente. “As observações ordenadas [...] são chamadas estatísticas de ordem” (BUSSAB e MORETTIN, 2010, p. 36).

Tomando um conjunto finito de valores como $x_1,\ x_2,\ x_3,\ \cdots,\ x_n;$ vamos denotar por x_i o i-ésimo valor de um rol quando os dados forem organizados em ordem crescente de forma que: $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_{n-1} \leq x_n$ e $i \in \{1, 2, 3, \cdots, n\}$.

Exemplo: Durante a correção de uma prova de matemática no valor de 10,0 pontos, o professor observou que as cinco primeiras notas (hipotéticas) foram 7,3, 6,4, 6,0, 3,5 e 6,0.

Colocando as notas em rol temos: 3,5, 6,0, 6,0, 6,4 e 7,3.

4.1.7 Classes

As classes são faixas de valores, ou valores individuais, ordenados na primeira coluna de uma tabela.

Exemplo: Usando as notas das provas do exemplo anterior, temos quatro classes:

Tabela 1 – Notas (hipotéticas) de cinco provas corrigidas

NOTAS	---	---
3,5	---	---
6,0	---	---
6,4	---	---
7,3	---	---

Fonte: Elaborado pelo autor (2020)

4.1.8 Amplitude da amostra

A amplitude da amostra é a diferença entre o maior valor (x_n) e o menor valor (x_1), observados em uma amostra.

Exemplo: Ainda tomando por base, as notas das provas do exemplo anterior, calcula-se a amplitude da amostra pela diferença entre a maior e menor notas

$$\text{amplitude da amostra} = x_5 - x_1$$

$$\text{amplitude da amostra} = 7{,}3 - 3{,}5$$

$$\text{amplitude da amostra} = 3{,}8.$$

4.1.9 Distribuição de frequência

Uma distribuição de frequência, é uma tabela que lista, em ordem, os dados de uma amostra juntamente com suas correspondentes frequências. Para Farber e Larson (2015, p. 37)

> Uma distribuição de frequência é uma tabela que mostra classes ou intervalos de valores com a contagem do número de ocorrências em cada classe ou intervalo. A frequência f de uma classe é o número de ocorrências de dados na classe.

A tabulação de dados brutos em uma distribuição de frequência, tem por objetivo a identificação da natureza comportamental destes dados para uma posterior descrição, exploração e comparação. "Quando se estuda uma variável, o maior interesse do pesquisador é conhecer o comportamento dessa variável, analisando a ocorrência de suas possíveis realizações" (BUSSAB e MORETTIN, 2010, p. 11).

As frequências são classificadas em:

1ª) frequência absoluta (f_a) que é a quantidade de vezes que um dado aparece na classe;

2ª) frequência absoluta acumulada (f_{aa}) que é a soma da frequência absoluta da classe com todas as frequências absolutas das classes que a antecedem, isto é:

$$f_{aa}^{n} = f_{a1} + f_{a2} + f_{a3} + \cdots + f_{an} = \sum_{i=1}^{n} f_{ai} \tag{4.1}$$

3ª) frequência relativa (f_r) que é a razão entre a frequência absoluta (f_a) da classe e a frequência absoluta acumulada (f_{aa});

4ª) frequência relativa acumulada (f_{ra}) que é a soma da frequência relativa da classe com todas as frequências relativas das classes que a antecedem, isto é:

$$f_{ra}^{n} = f_{r1} + f_{r2} + f_{r3} + \cdots + f_{rn} = \sum_{i=1}^{n} f_{ri} \tag{4.2}$$

Exemplo: O tempo de utilização de caixas eletrônicos nas agências bancárias por um cliente, depende de quem está usando e de quais operações serão efetuadas. Na tabela abaixo temos o registro, em rol, dos tempos (hipotéticos) que 30 clientes gastaram para efetuar suas transações.

Quadro 1 – Tempos (hipotéticos) de utilização de caixas eletrônicos

Tempo aproximado em minutos														
2,8	2,8	3,0	3,0	3,0	3,0	3,0	3,6	3,6	3,6	3,8	4,1	4,1	4,1	4,1
4,1	4,1	4,1	4,1	4,5	4,5	4,5	4,8	4,8	4,8	4,9	5,2	5,2	5,2	7,7

Fonte: Elaborado pelo autor (2020)

Construindo uma tabela de distribuição de frequências cujas classes são os tempos, temos:

Tabela 2 – Tabela de frequência para o tempo utilização de caixas eletrônicos

TEMPOS	f_a	f_{aa}	f_r	f_{ra}
2,8	2	2	0,0667	0,0667
3,0	5	7	0,1667	0,2334
3,6	3	10	0,1000	0,3334
3,8	1	11	0,0333	0,3667
4,1	8	19	0,2667	0,6334
4,5	3	22	0,1000	0,7334
4,8	3	25	0,1000	0,8334
4,9	1	26	0,0333	0,8667
5,2	3	29	0,1000	0,9667
7,7	1	30	0,0333	1,0000

Fonte: Elaborado pelo autor (2020)

Observamos que a frequência relativa, usualmente, pode ser interpretada na forma percentual. Assim, para o tempo de 5,2 minutos teríamos $\mathrm{f_r} = 0{,}1000 = 10{,}0\%$ dos valores registrados.

Quando trabalhamos com variáveis contínuas, podemos usar intervalos numéricos determinando faixas de valores para as classes de uma tabela. Aqui, o objetivo é sintetizar os dados.

Exemplo: Usando os tempos do quadro 1, vamos trabalhar as classes em 6 intervalos de classe e justificaremos esta escolha posteriormente. Primeiro determinamos a amplitude de cada intervalo dividindo a amplitude da amostra por 6.

$$\frac{\mathrm{x}_{30} - \mathrm{x}_1}{6} = \frac{7{,}7 - 2{,}8}{6} = 0{,}816666\ldots \approx 0{,}82$$

Observamos que a amplitude do intervalo sempre será ajustada para um valor maior e conveniente, com o objetivo de contemplar a frequência dos últimos dados listados.

Usaremos a notação $\mathbf{a} \vdash \mathbf{b}$ para designar o intervalo numérico que inclui o valor de **a** mas não inclui o valor de **b**. Na tabela 4, temos que $\mathbf{a} = 2{,}80$ e $\mathbf{b} = \mathbf{a} + 0{,}82 = 3{,}62$.

Tabela 3 – Tabela de frequência em intervalos de classe

TEMPOS	f_a	f_{aa}	f_r	f_{ra}
$2{,}80 \vdash 3{,}62$	10	10	0,3333	0,3333
$3{,}62 \vdash 4{,}44$	9	19	0,3000	0,6333
$4{,}44 \vdash 5{,}26$	10	29	0,3333	0,9667
$5{,}26 \vdash 6{,}08$	0	29	0,0000	0,0000
$6{,}08 \vdash 6{,}90$	0	29	0,0000	0,0000
$6{,}90 \vdash 7{,}72$	1	30	0,0333	1,0000

Fonte: Elaborado pelo autor (2020)

Como a proposta de trabalho tem a Educação Básica como público-alvo, é interessante limitar a quantidade de intervalos a serem construídos no intuito de orientar e estabelecer um padrão das tabelas que serão criadas. Para determinar a quantidade de intervalos a serem usados nas classes, não tem uma regra geral. Segundo Bussab e Morettin (2010, p. 13)

> A escolha dos intervalos é arbitrária e a familiaridade do pesquisador com os dados é que lhe indicará quantas e quais classes (intervalos) devem ser usadas. Entretanto, deve-se observar que, com um pequeno número de classes, perde-se informação, e com um número grande de classes, o objetivo de resumir os dados fica prejudicado. Estes dois extremos têm a ver, também, com o grau de suavidade da representação gráfica dos dados, [...]. Normalmente, sugere-se o uso de 5 a 15 classes com a mesma amplitude.

Vamos adotar a regra de Sturges para calcular a quantidade **k** de intervalos. Nesta regra temos:

$$\mathbf{k} = 1 + 3{,}22\,.\log \mathbf{n} \tag{4.3}$$

onde **n** é a quantidade de dados da amostra e $\mathbf{k} \in \{5, 6, 7, \cdots, 15\}$**.**

No exemplo anterior adotou-se convenientemente 6 intervalos de classe, os quais justificam-se agora. Note que para o conjunto das observações temos 30 tempos em minutos. Assim, para $\mathrm{n} = 30$

$$\mathbf{k} = 1 + 3{,}22\,.\log \mathbf{30} \Leftrightarrow \mathbf{k} \approx 5{,}76 \Leftrightarrow \mathbf{k} = 6{,}0.$$

4.1.10 Ponto médio de um intervalo de classe

É o valor central no intervalo de classe.

Para o cálculo do ponto médio da i-ésima classe (m_i), procederemos assim:

$$m_i = \frac{v_I + v_S}{2} \quad \text{com } i \in \{1, 2, 3, \cdots, k-1, k\} \tag{4.4}$$

$$\text{onde } \begin{cases} v_I \text{ é o valor do limite inferior da classe} \\ v_S \text{ é o valor do limite superior da classe} \end{cases}$$

Exemplo: Usando a tabela 3 como referência, note que para o primeiro intervalo de classe $(i = 1)$ temos $v_I = 2{,}80$ e $v_S = 3{,}62$:

$$m_1 = \frac{v_I + v_S}{2} = \frac{2{,}80 + 3{,}62}{2} \Leftrightarrow m_1 = 3{,}21$$

4.1.11 Tabela de contingência

Uma tabela de contingência faz distribuições conjuntas de duas variáveis (análise bivariada).

Exemplo: Dados do Datasus (2020), para a consolidação do Sistema de Informações Sobre Nascidos Vivos que está em vigor desde 2011, mostram os registros dos nascimentos no ano de 2018 para a região sudeste. Observe a tabela 4.

Tabela 4 – Nascidos vivos na região sudeste (2018)

Nascimento por ocorrência por Sexo e Unidade da Federação

Região Sudeste

Período: 2018

Sexo	MG	ES	RJ	SP	Total
Masculino	134.420	29.003	112.424	311.425	587.272
Feminino	128.615	27.325	108.028	296.191	560.159
Total	263.035	56.328	220.452	607.616	1.147.431

Fonte: Adaptado https://datasus.saude.gov.br/informacoes-de-saude-tabnet/
Acesso em: 04/04/2020

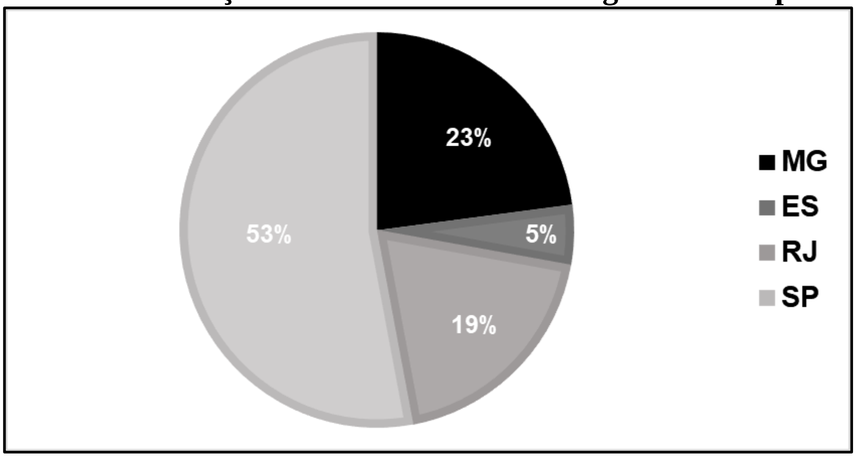
Nota-se que este tipo de tabela é muito útil quando há o interesse de estudar duas variáveis para um enriquecimento de detalhes do conjunto de observações.

4.1.12 Gráfico de setores

Um gráfico de setores é formado a partir da divisão de um círculo em setores cujas áreas serão proporcionais às frequências que cada variável apresentar. Pesquisadores utilizam com certa regularidade esse tipo de gráfico para retratar dados qualitativos.

Exemplo: Dados do Datasus (2020) apontam que em 2018 foram registrados 1.147.431 nascidos-vivos na região sudeste, dos quais 587.272 eram do sexo masculino e 560.159 do sexo feminino. Observe o Gráfico 2 que faz uma distribuição dos nascimentos por estado segundo a tabela 4.

Gráfico 2 – Distribuição dos nascidos-vivos na região sudeste por estado



Fonte: Adaptado https://datasus.saude.gov.br/informacoes-de-saude-tabnet/
Acesso em: 04/04/2020

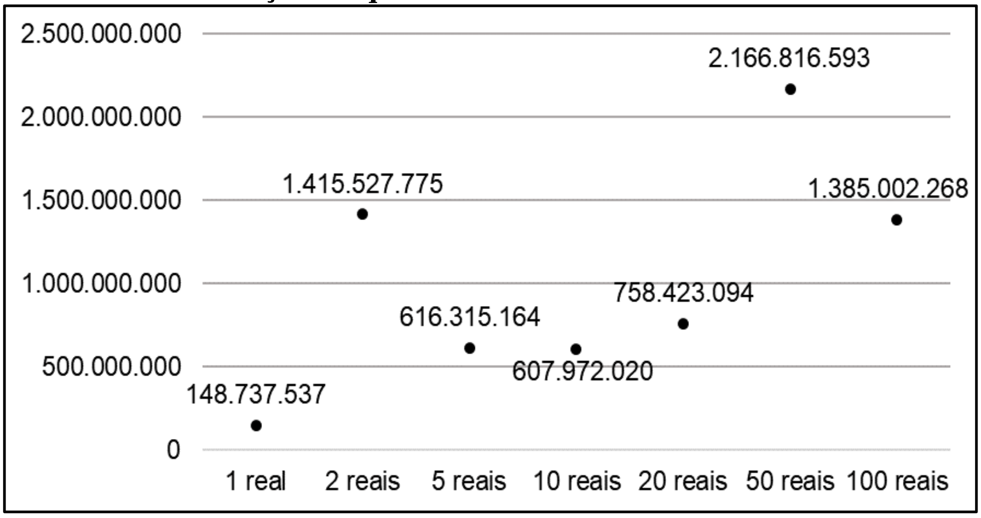
4.1.13 Diagrama de dispersão

O diagrama de dispersão é um gráfico de pontos distribuídos em um plano cartesiano de eixos reais, cujos pares de dados (x, y) estão de acordo com sua abscissa e ordenada.

Exemplo: O Banco Central do Brasil controla a quantidade de cédulas e moedas metálicas do padrão monetário Real que estão em circulação no Brasil durante os dias úteis de um ano.

O Gráfico 3 mostra uma distribuição dos valores que estavam em poder do público e da rede bancária no primeiro dia útil de 2020.

Gráfico 3 – Distribuição da quantidade de cédulas e moedas em 02/01/2020



Fonte: Adaptado https://www3.bcb.gov.br/mecpublico/circulante
Acesso em: 08/04/2020

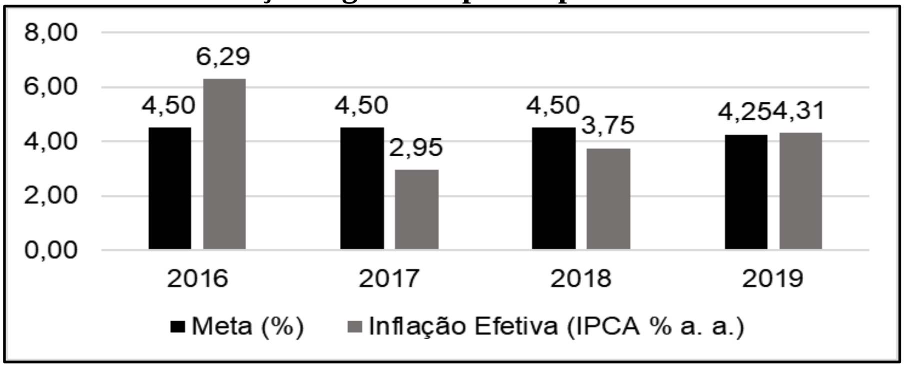
4.1.14 Gráfico de série temporal

Um gráfico de série temporal distribui e organiza os dados coletados de uma amostra conforme as datas ou períodos diferentes em que cada um ocorreu.

Exemplo 1: O regime de metas para a inflação tem sido bem sucedido no Brasil. O sistema tem possibilitado que a inflação fique sob controle, em níveis relativamente baixos. Desde a adoção do regime em 1999, a inflação tem situado dentro do intervalo de tolerância na maioria dos anos-calendário.

O Gráfico 4 mostra os registros da meta prevista e da inflação efetiva para os anos de 2016 a 2019.

Gráfico 4 – Inflação registrada para o período de 2016 a 2019



Fonte: Adaptado https://www.bcb.gov.br/controleinflacao/historicometas
Acesso em: 02/04/2020

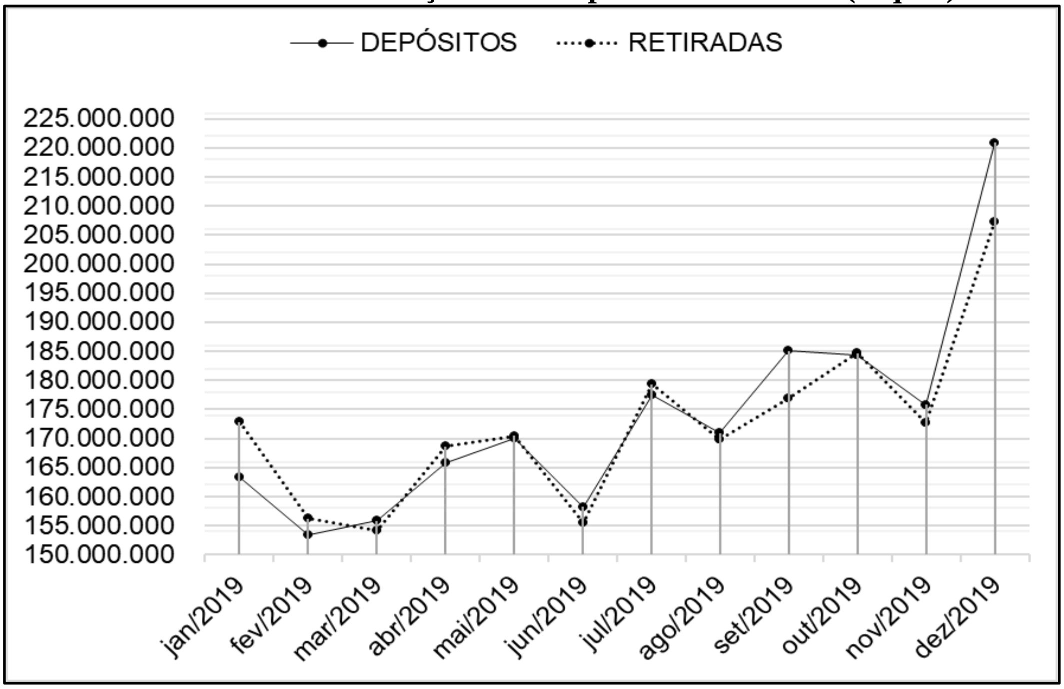
Exemplo 2: Economistas sempre alertam que a poupança não é um investimento financeiro com um bom retorno. Mas sabe-se que ela é na prática, a aplicação mais adotada pela população economicamente ativa no Brasil. Observe a tabela 5 e o Gráfico 5 que mostram como variaram os depósitos e as retiradas (saques) na caderneta de poupança durante o ano de 2019.

Tabela 5 – Movimentação na caderneta de poupança durante o ano de 2019

TRANSAÇÕES REGISTRADAS			
mês / ano	**depósitos (A)**	**retiradas (B)**	**capitação líquida (A – B)**
jan / 2019	163.409.961	172.815.706	-9.405.745
fev / 2019	153.362.874	156.197.008	-2.834.134
mar / 2019	155.820.184	154.253.027	1.567.157
abr / 2019	165.821.579	168.647.311	-2.825.732
mai / 2019	170.025.741	170.489.617	-463.876
jun / 2019	158.228.614	155.458.436	2.770.178
jul / 2019	177.499.813	179.335.929	-1.836.116
ago / 2019	170.952.894	169.923.427	1.029.467
set / 2019	185.044.915	176.952.378	8.092.537
out / 2019	184.316.496	184.667.104	-350.608
nov / 2019	175.757.990	172.708.866	3.049.124
dez / 2019	220.879.028	207.281.705	13.597.323
TOTAL	2.081.120.089	2.068.730.514	12.389.575

Fonte: Adaptado https://www.bcb.gov.br/estabilidadefinanceira/relatoriopoupanca
Acesso em: 02/04/2020

Gráfico 5 – Movimentação entre depósitos e retiradas (saques)



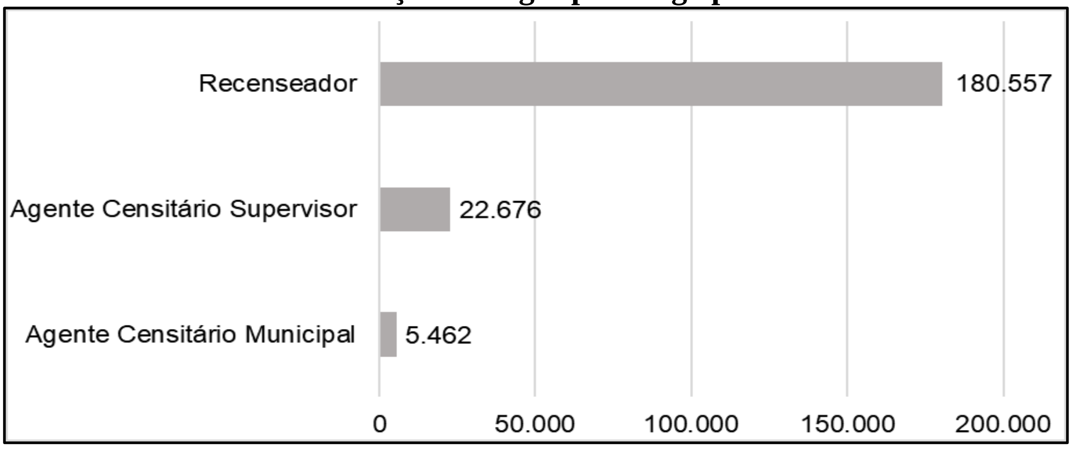
Fonte: https://www.bcb.gov.br/estabilidadefinanceira/relatoriopoupanca
Acesso em: 02/04/2020

É notório que a tabela coloca uma riqueza de detalhes que respondem às indagações que possam surgir enquanto o gráfico simplifica visualmente a movimentação, entre depósitos e retiradas, mostrando as variações em cada período.

4.1.15 Gráfico de pareto

O gráfico de pareto é usado em análises de dados qualitativos com barras, na horizontal ou na vertical, dispostas em ordem pela frequência que facilita a visualização dos dados.
Exemplo: Para a realização do Censo Demográfico 2020, serão contratados temporariamente 208.695 recenseadores através dos Processos Seletivos Simplificados, cujas inscrições serão realizadas exclusivamente no site do Centro Brasileiro de Pesquisa em Avaliação e Seleção e de Promoção de Eventos (Cebraspe), banca organizadora dos processos seletivos. As vagas são para os cargos de recenseador, agente censitário municipal e agente censitário supervisor.
O Gráfico 6 mostra a distribuição das vagas segundo o cargo.

Gráfico 6 – Distribuição de vagas por cargo para o Censo 2020



Fonte: Adaptado https://censo2020.ibge.gov.br/2963-c2020-censo-2020/c2020-hotsite/27158-inscricoes-nos-processos-seletivos-para-agente-censitario-e-recenseador-terminam-na-proxima-terca.html
Acesso em: 02/04/2020

4.1.16 Histograma

Segundo Triola (2008, p. 41)

> Um histograma é um gráfico de barras no qual a escala horizontal representa classes de valores de dados e a escala vertical representa frequências. As

> alturas das barras correspondem aos valores das frequências, e as barras são desenhadas adjacentes umas às outras (sem separação).

Aqui, observamos que um histograma é basicamente a versão gráfica de uma tabela de frequência com as classes em intervalos.

Exemplo: A cotação do dólar é feita de acordo com a taxa de câmbio que expressa a relação econômica entre Brasil e Estados Unidos. É o câmbio que define quantos reais são necessários para comprar um dólar. Sua taxa é flexível porque depende de muitos fatores que influenciam as transações comerciais como política, insumos e fenômenos naturais. Na tabela 6, temos o registro da cotação do dólar para o primeiro trimestre de 2020.

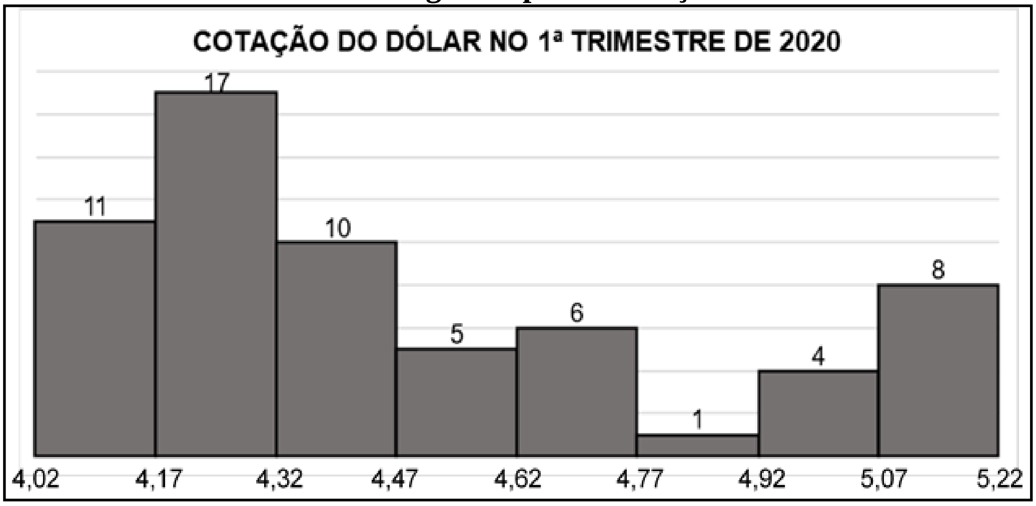
Tabela 6 – Cotação do dólar comercial no primeiro trimestre de 2020

DATA	COTAÇÃO	DATA	COTAÇÃO
02 / jan	R$ 4,02	14 / fev	R$ 4,32
03 / jan	R$ 4,05	17 / fev	R$ 4,32
06 / jan	R$ 4,05	18 / fev	R$ 4,35
07 / jan	R$ 4,08	19 / fev	R$ 4,37
08 / jan	R$ 4,07	20 / fev	R$ 4,39
09 / jan	R$ 4,07	21 / fev	R$ 4,39
10 / jan	R$ 4,07	26 / fev	R$ 4,44
13 / jan	R$ 4,13	27 / fev	R$ 4,48
14 / jan	R$ 4,14	28 / fev	R$ 4,50
15 / jan	R$ 4,16	02 / mar	R$ 4,49
16 / jan	R$ 4,17	03 / mar	R$ 4,49
17 / jan	R$ 4,18	04 / mar	R$ 4,53
20 / jan	R$ 4,18	05 / mar	R$ 4,62
21 / jan	R$ 4,20	06 / mar	R$ 4,65
22 / jan	R$ 4,19	09 / mar	R$ 4,74
23 / jan	R$ 4,17	10 / mar	R$ 4,67
24 / jan	R$ 4,18	11 / mar	R$ 4,67
27 / jan	R$ 4,22	12 / mar	R$ 4,88
28 / jan	R$ 4,21	13 / mar	R$ 4,74
29 / jan	R$ 4,20	16 / mar	R$ 4,95
30 / jan	R$ 4,25	17 / mar	R$ 5,05
31 / jan	R$ 4,27	18 / mar	R$ 5,11
03 / fev	R$ 4,25	19 / mar	R$ 5,14
04 / fev	R$ 4,24	20 / mar	R$ 5,02
05 / fev	R$ 4,24	23 / mar	R$ 5,08
06 / fev	R$ 4,25	24 / mar	R$ 5,07
07 / fev	R$ 4,31	25 / mar	R$ 5,07
10 / fev	R$ 4,32	26 / mar	R$ 5,00
11 / fev	R$ 4,31	27 / mar	R$ 5,11
12 / fev	R$ 4,34	30 / mar	R$ 5,16
13 / fev	R$ 4,34	31 / mar	R$ 5,20

Fonte: Adaptado http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38590&module=M
Acesso em: 02/04/2020

Para uma análise mais rápida, constrói-se o Gráfico 7 como um histograma que mostra oito classes, em intervalos de preços, e as quantidades (frequências) de valores observados (cotados) em cada classe.

Gráfico 7 – Histograma para a cotação do dólar



Fonte: Adaptado http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38590&module=M
Acesso em: 02/04/2020

Nota-se que a tabela 6, cumpre sua função de transmitir as informações de forma clara, precisa e detalhada sobre valores e datas, porém, tem a desvantagem de ser muito extensa, o que pode levar a alguma confusão durante a leitura dos dados. Já o histograma apresentado no Gráfico 7, não tem a mesma riqueza de detalhes, mas, também transmite as informações de forma clara e precisa simplificando a leitura dos dados pelos intervalos e frequências que não deixam margem para alguma interpretação equivocada.

4.1.17 Polígono de frequência

O polígono de frequência é um gráfico de linha contínua obtido pela junção de segmentos de reta cujas extremidades são pontos coordenados. Estes pontos têm por abscissa, o ponto médio do intervalo de classe, e por ordenada, a frequência simples da classe. As extremidades de início e fim, sempre terão frequência zero.

Exemplo: Usando o histograma do exemplo 4.15.1, vamos calcular o ponto médio do primeiro intervalo de classe (4.02, 4.17). Pela definição 4.10, temos:

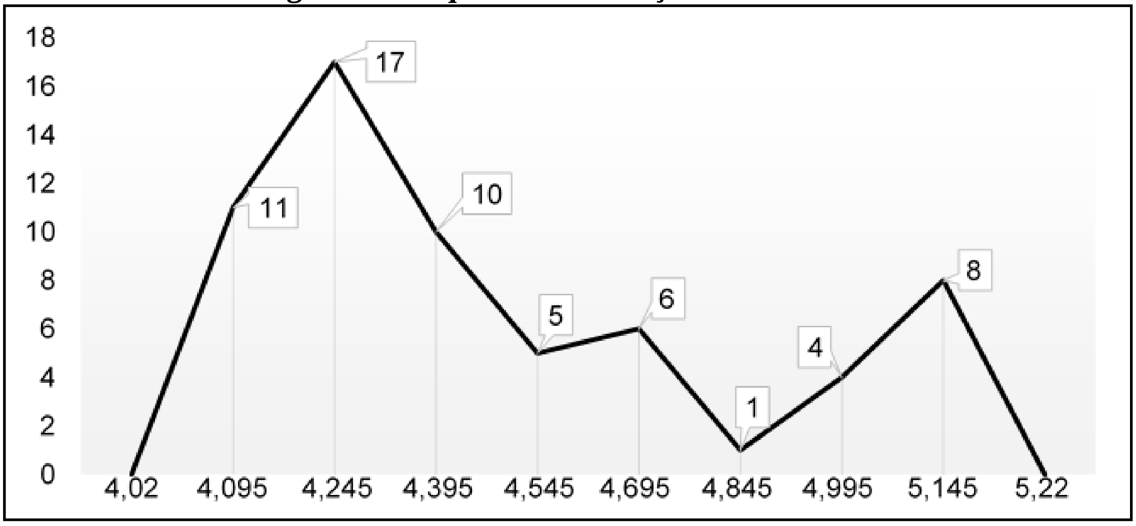
$$\mathrm{m}_1 = \frac{\mathrm{v_I} + \mathrm{v_S}}{2} = \frac{4{,}02 + 4{,}17}{2} \Leftrightarrow \mathrm{m}_1 = 4.095$$

onde obtemos o ponto (4.095, 11.0).

Notemos que este não será o primeiro ponto do polígono de frequência pois, o primeiro e último pontos terão ordenada zero. Uma sugestão para o primeiro ponto pode ser o par ordenado (4.02, 0.0) e seguindo para os posteriores obteremos:

(4.02, 0.0), (4.095, 11.0), (4.245, 17.0), (4.395, 10.0), ... , (5.22, 0.0).

Gráfico 8 – Polígono de frequência da cotação do dólar - 1º trim. de 2020



Fonte: Adaptado http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38590&module=M
Acesso em: 02/04/2020

Até o momento, apresentamos tabelas que permitem organizar os dados de uma amostra fazendo uma distribuição das frequências e alguns tipos de gráficos que resumem e apresentam a distribuição dos dados visualmente. Todavia, precisamos também de parâmetros que mostrem, através de uma medida numérica, algumas características da amostra.

Como colocado por Farias (2020, p. 21)

> A redução dos dados através de tabelas de frequências ou gráficos é um dos procedimentos disponíveis para se ilustrar o comportamento de um conjunto de dados. No entanto, muitas vezes, queremos resumir ainda mais esses dados, apresentando valores únicos que descrevam suas principais características.

Assim, continuaremos a descrição, exploração e comparação de dados através de medidas de centro (média, moda e mediana), medidas de posições relativas (separatrizes, distribuição simétrica e assimétrica) e medidas de dispersões (desvio médio, variância e desvio padrão).

4.1.18 Média

A média $\overline{x}$ (lê-se: x barra) entre n observações $x_1, x_2, x_3, \cdots, x_{n-1}, x_n$; é dada pela razão entre a soma destes valores e n.

$$\bar{x} = \frac{x_1+x_2+x_3+\cdots+x_{n-1}+x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (4.5)$$

Exemplo: Vamos calcular a média de nascidos vivos por estado na região sudestes pelos dados da tabela 4.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_{n-1} + x_n}{n}$$

$$\bar{x} = \frac{263.035 + 56.328 + 220.542 + 607.616}{4}$$

$$\bar{x} = \frac{1.147.431}{4} = 286.857{,}75$$

Onde temos um valor médio de 286.857,75 nascidos vivos por estado.

Para o cálculo da média de dados distribuídos em intervalos de classes, consideremos o ponto médio m_i, como na definição 4.1.10, e a frequência absoluta f_{ai} do i-ésimo intervalo. Assim, determinamos a média por:

$$\bar{x} = \frac{f_{a1}.m_1+f_{a2}.m_2+f_{a3}.m_3+\cdots+f_{an}.m_n}{f_{a1}+f_{a2}+f_{a3}+\cdots+f_{an}} = \frac{1}{\sum_{i=1}^{n} f_{ai}} \sum_{i=1}^{n} f_{ai}.m_i \quad (4.6)$$

Exemplo: Vamos calcular a média dos valores cotados do dólar no primeiro trimestre de 2020 pelo polígono de frequência (Gráfico 8).

$$\bar{x} = \frac{f_{a1}.m_1 + f_{a2}.m_2 + f_{a3}.m_3 + \cdots + f_{an}.m_n}{f_{a1} + f_{a2} + f_{a3} + \cdots + f_{an}}$$

$$\bar{x} = \frac{11\ .\ 4{,}095 + 17\ .\ 4{,}245 + 10\ .4{,}395 + 5\ .\ 4{,}545 + 6\ .4{,}695 + 1\ .4{,}845 + 4\ .4{,}995 + 8\ .5{,}145}{11 + 17 + 10 + 5 + 6 + 1 + 4 + 8}$$

$$\bar{x} = \frac{278{,}04}{62} = 4{,}485$$

Onde temos um valor médio de R$ 4,485.

Aqui, vale ressaltar dois pontos:

- a média para um conjunto de dados deve ser, prioritariamente, calculada a partir dos dados brutos;
- os valores das medidas de centro, posição relativa e de dispersão para um conjunto de dados observados; podem ser influenciados por valores discrepantes (*outliers*), isto é, valores que se destoam da maior parte dos dados observados por serem muito grandes ou muito pequenos.

Observa-se que alguns *outliers* podem ocorrer por erros no registro dos dados. Ao fazer uma distribuição de um conjunto de dados que possui um ou mais *outliers*, surgem lacunas que interferem nas análises e levam a conclusões que podem ser falhas.

Colocando de forma mais direta, voltemos ao quadro 1 e à tabela 3 onde temos um *outlier* cujo valor é 7,7 minutos. Este tempo, coloca uma lacuna de dois intervalos sem dados, como na tabela 3, quando comparado com o último, levando a uma indicação de possível erro na coleta dos dados.

4.1.19 Moda

A moda x^* (lê-se: x asterisco) de um conjunto de n dados observados é o valor que mais vezes repete, ou seja, o valor de maior frequência absoluta.

Exemplo 1: Considere um conjunto de cinco notas hipotéticas, em rol, para cinco alunos distintos numa prova de matemática: 4, 6, 6, 6 e 9. Como a nota 6 tem frequência absoluta três e nenhuma outra nota tem frequência maior ou igual, temos que esta nota é a moda do conjunto de notas observadas onde $x^* = 6$.

Segundo Triolla (2008), vale ressaltar a possibilidade de uma distribuição de valores não ter moda, denominada como distribuição amodal, ou seja, todos os valores têm a mesma frequência absoluta. Ter uma moda, distribuição modal, significa que um único valor tem a maior frequência absoluta. Já, possuir duas modas, distribuição bimodal, dois valores empatam na maior frequência absoluta. Caso haja três modas ou mais, classifica-se a distribuição como multimodal.

Exemplo 2: Novamente, observando os dados da tabela 6 onde temos a cotação do dólar comercial no primeiro trimestre de 2020, notamos que os valores R$ 4.07, R$ 4.18, R$ 4.25 e R$ 4.32; tem a mesma frequência absoluta três e nenhuma outra cotação tem frequência igual ou maior. Logo, todas estas cotações são as modas, onde classificamos o conjunto como multimodal.

4.1.20 Mediana

A mediana $\tilde{x}$ (lê-se: x til) é o valor que ocupa a posição central de um conjunto de dados observados, quando estes estão em rol.

Conforme Farias (2020), a mediana $\tilde{x}$ é o valor pertencente a um conjunto de dados observados em ordem crescente tal que, 50% dos dados são menores que $\tilde{x}$ e os outros 50% são maiores que ele.

Para fins de cálculo, consideremos um conjunto de dados observados com n elementos em rol: $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_{n-1} \leq x_n$.

$$\text{Se n é ímpar: } \tilde{x} = x_{\frac{n+1}{2}} \tag{4.7}$$

$$\text{Se n é par: } \tilde{x} = \frac{x_{\frac{n}{2}} + x_{\frac{n}{2}+1}}{2} \tag{4.8}$$

Exemplo 1: Considere um conjunto de cinco notas hipotéticas, em rol, para cinco alunos distintos numa prova de matemática: 4, 6, 6, 7 e 9. Como n = 5, temos: $\tilde{x} = x_{\frac{5+1}{2}} = x_3 = 6$.

Exemplo 2: Considere um conjunto de seis notas hipotéticas, em rol, para seis alunos distintos numa prova de matemática: 4, 6, 6, 7, 9 e 10. Como n = 6, temos: $\tilde{x} = \frac{x_{\frac{6}{2}} + x_{\frac{6}{2}+1}}{2} = \frac{x_3 + x_4}{2} = 6{,}5$.

Quando for preciso calcular a mediana em um histograma ou uma tabela de intervalos de classes, seguiremos os passos:

1º) calcular a frequência absoluta acumulada da última classe obtendo o total de valores observados da amostra;

2º) dividir este total por 2;

3º) pelo valor obtido no segundo passo, identificamos a classe que contém a mediana, ou seja, o intervalo de classe mediana;

4º) aplicar a definição 4.1.10, calculando o ponto médio deste intervalo

$$\tilde{x} = m_i = \frac{v_I + v_S}{2}\text{, onde} \begin{cases} v_I \text{ é o valor do limite inferior da classe} \\ v_S \text{ é o valor do limite superior da classe} \end{cases}$$

Exemplo 3: A partir do histograma, Gráfico 7, vamos calcular a mediana da distribuição feita com a cotação do dólar.

1º) $\sum_{i=1}^{8} f_{ai} = f_{a1} + f_{a2} + f_{a3} + f_{a4} + f_{a5} + f_{a6} + f_{a7} + f_{a8} = 62;$

2º) $\frac{\sum_{i=1}^{8} f_{ai}}{2} = \frac{62}{2} = 31;$

3º) por este valor obtido, identificamos que a classe que contém a mediana é o intervalo $4{,}32 \vdash 4{,}47$;

4º) aplicando a definição 4.1.10, temos:

$$\tilde{x} = \frac{v_I + v_S}{2} = \frac{4{,}32 + 4{,}47}{2} = 4{,}395$$

4.1.21 Separatrizes

"A Separatriz de ordem p é um valor tal que pelo menos p% dos dados são menores que ele e pelo menos (1 – p) % são maiores" (FARIAS, 2020, p. 26).

Por esta definição, observamos que a mediana é uma separatriz. As separatrizes mais utilizadas são: mediana, quartis, decis e percentis. Aqui, usaremos somente a definição dos quartis para dar sequência ao trabalho com o objetivo de delimitar as definições e análises estatísticas, posto que o foco são estudantes da Educação Básica.

4.1.22 Quartis

Toma-se um conjunto de dados observados em rol. Chamaremos de quartis a três valores que dividem este conjunto de dados em quatro partes iguais.

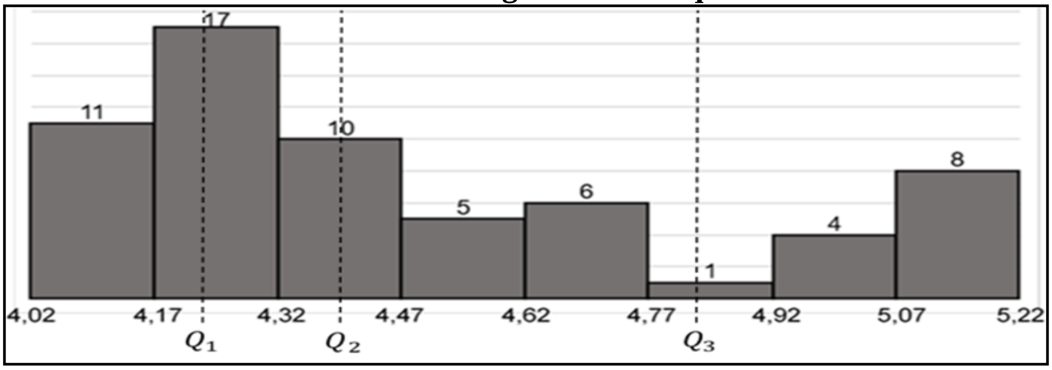
Desse modo, os três quartis serão:

- primeiro quartil (Q_1) é o valor que coloca pelo menos 25% das observações abaixo dele e pelo menos 75% das outras acima dele;
- segundo quartil (Q_2) é o valor que coloca pelo menos 50% das observações abaixo dele e pelo menos 50% das outras acima dele;
- terceiro quartil (Q_3) é o valor que coloca pelo menos 75% das observações abaixo dele e pelo menos 25% das outras acima dele.

Nota-se que Q_2 é a mediana do conjunto de dados, ou seja, $Q_2 = \tilde{x}$.

Exemplo: Usando novamente o histograma que distribui a cotação do dólar no 1ª trimestre de 2020, Gráfico 7, vamos determinar os valores de Q_1, Q_2 e Q_3, Gráfico 9.

Gráfico 9 – Histograma e seus quartis



Fonte: Adaptado http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38590&module=M
Acesso em: 02/04/2020

I) Pelo exemplo da definição 4.1.20 que é mediana ou segundo quartil: $Q_2 = \tilde{x} = 4{,}395$. Analisando temos que 50% da distribuição dos dados têm valores menores que 4,395 e os outros 50% têm valores maiores.

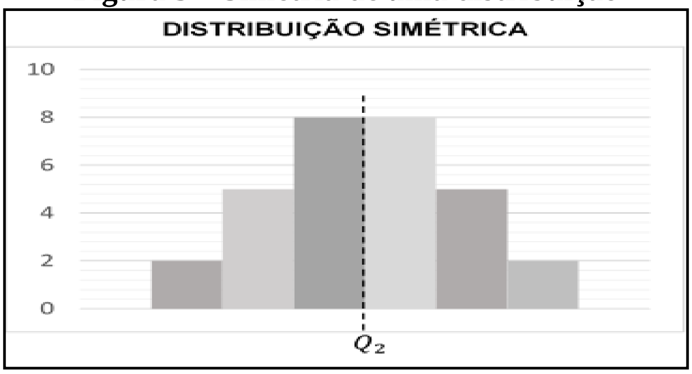
II) $Q_1 = \frac{4{,}02+4{,}395}{2} = 4{,}208$. Analisando temos que 25% da distribuição dos dados têm valores menores que 4,208 e os outros 75% têm valores maiores.

III) $Q_3 = \frac{4{,}395+5{,}22}{2} = 4{,}808$. Analisando temos que 75% da distribuição dos dados têm valores menores que 4,808 e os outros 25% têm valores maiores.

4.1.23 Simetria de uma distribuição

Segundo Farias (2020), a distribuição feita em um histograma, diagrama de pontos ou gráfico de barras é simétrica, quando os lados direito e esquerdo são, aproximadamente, a imagem espelhada um do outro em relação a Q_2.

Figura 3 – Simetria de uma distribuição

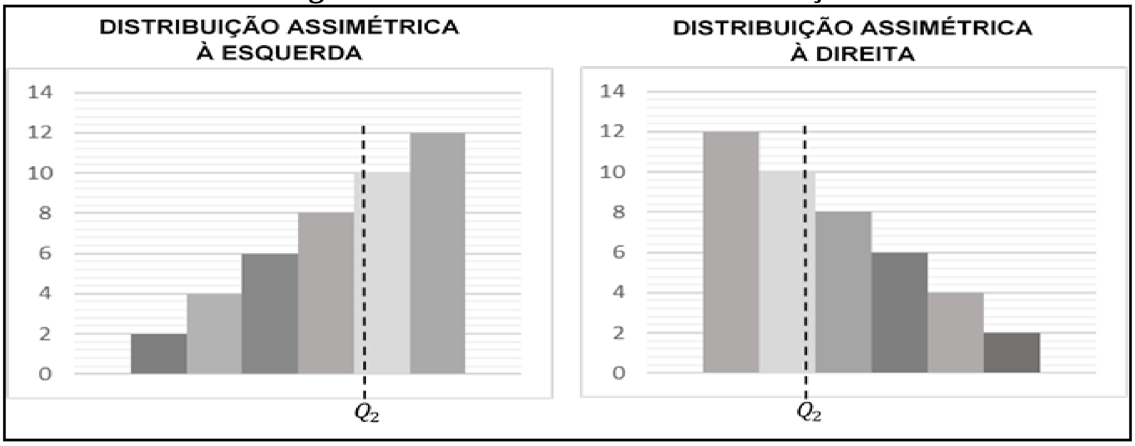


Fonte: Elaborado pelo autor (2020)

4.1.24 Assimetria de uma distribuição

"Uma distribuição é assimétrica à direita se a cauda direita do histograma se estende muito mais do que a cauda esquerda. Ela é assimétrica à esquerda se a cauda esquerda do histograma se estende muito mais do que a cauda direita" (FARIAS, 2020, p. 40).

Figura 4 – Assimetria de uma distribuição



Fonte: Elaborado pelo autor (2020)

Pode-se fazer uma análise da distribuição de um conjunto de dados usando os quartis. Como entre Q_1 e Q_2 e entre Q_2 e Q_3 temos 25% dos dados, as diferenças $(Q_2 - Q_1)$ e $(Q_3 - Q_2)$ darão uma informação sobre a simetria ou assimetria da distribuição, onde observa-se que:

- se $(Q_3 - Q_2) - (Q_2 - Q_1) > 0$ temos uma assimetria positiva pois, a distribuição será assimétrica à direita;
- se $(Q_3 - Q_2) - (Q_2 - Q_1) < 0$ temos uma assimetria negativa pois, a distribuição será assimétrica à esquerda;
- se $(Q_3 - Q_2) - (Q_2 - Q_1) = 0$ temos uma simetria (ou simetria nula).

4.1.25 Coeficiente de assimetria de Bowley

Segundo Farias (2020), este coeficiente é definido como:

$$B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} \qquad (4.9)$$

Após o cálculo desse coeficiente, notaremos que:

- quanto mais próximos forem Q_1 e Q_2, a distribuição será assimétrica à direita e B será um valor mais próximo de $+1$.
- quanto mais próximos forem Q_2 e Q_3, a distribuição será assimétrica à esquerda e B será um valor mais próximo de -1.
- quanto mais próximos forem os valores das diferenças $(Q_3 - Q_2)$ e $(Q_2 - Q_1)$, a distribuição tende a ser simétrica e B será um valor mais próximo de 0.

Exemplo: Dados os valores de $Q_1 = 4.208$, $Q_2 = 4.395$ e $Q_3 = 4.808$ pertencentes ao Histograma e seus Quartis no Gráfico 9, vamos analisar a distribuição da cotação do dólar no 1º trimestre de 2020, quanto a sua simetria ou assimetria.

I) pelos quartis:

$$(Q_3 - Q_2) - (Q_2 - Q_1) = (4{,}808 - 4{,}395) - (4{,}395 - 4{,}208) = 0{,}226 > 0$$

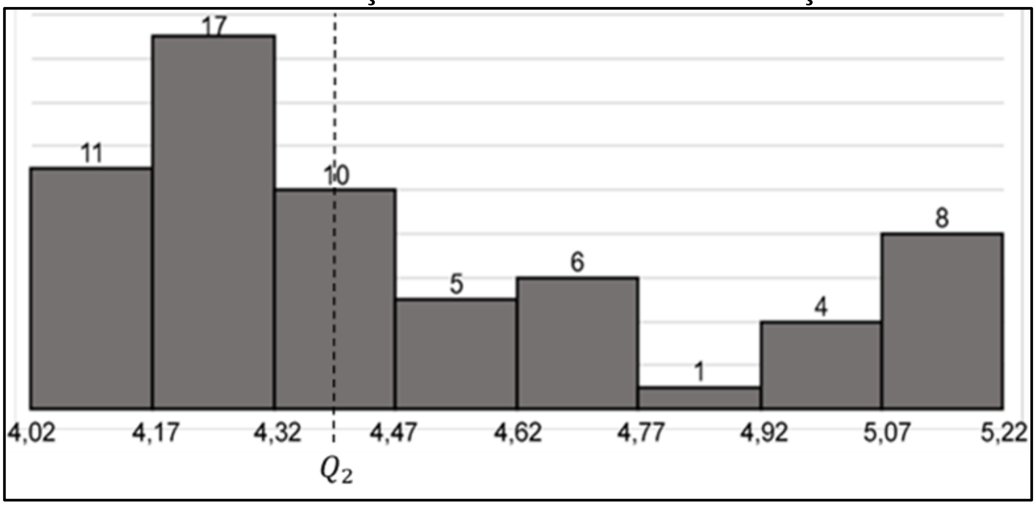
e conclui-se que a distribuição é assimétrica à direita, o que nos dá uma concentração (frequência) de valores maior na primeira metade do trimestre em comparação com a segunda metade (calda direita).

II) analisando agora, pelo coeficiente de Bowley, temos:

$$B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)} = \frac{(4{,}808 - 4{,}395) - (4{,}395 - 4{,}208)}{(4{,}808 - 4{,}208)} = 0{,}377$$

e conclui-se, como anteriormente, que Q_1 e Q_2 estão mais próximos resultando em uma distribuição assimétrica à direita, Gráfico 10.

Gráfico 10 – Distribuição assimétrica à direita da cotação do dólar



Fonte: Adaptado http://www.ipeadata.gov.br/ExibeSerie.aspx?serid=38590&module=M
Acesso em: 02/04/2020

De acordo com Bussab e Morettin (2010, p. 37, p. 38 e p. 39)

> O resumo de um conjunto de dados por uma única medida representativa de posição central esconde toda a informação sobre a variabilidade do conjunto de observações. [...]
> Notamos, então, a conveniência de serem criadas medidas que sumarizem a variabilidade de um conjunto de observações e que nos permita, por exemplo, comparar conjuntos diferentes de valores, [...], segundo algum critério estabelecido.
> Um critério frequentemente usado para tal fim é aquele que mede a dispersão dos dados em torno de sua média, e duas medidas são as mais usadas: desvio médio e variância. O princípio básico é analisar os desvios das observações em relação à média dessas observações. [...]
> Sendo a variância uma medida de dimensão igual ao quadrado da dimensão dos dados [...], ela pode causar problemas de interpretação. Costuma-se usar, então, o desvio padrão, que é definido como a raiz quadrada positiva da variância.

4.1.26 Desvio de um dado

O desvio de um dado é a diferença entre o valor de um dado x_i pertencente a um conjunto X de observações e o valor da média $\overline{x}$ dos dados deste conjunto: $(x_i - \overline{x})$.

Observam-se dois pontos importantes: o primeiro é que o desvio de um dado mostra a dispersão deste dado x_i em relação à média $\overline{x}$ e o segundo, é que a soma dos desvios de qualquer conjunto de dados sempre será igual a zero.

4.1.27 Desvio médio

O desvio médio de um conjunto X de observações (dados), é a média dos módulos dos desvios.

$$dm(X) = \frac{|x_1 - \overline{x}| + |x_2 - \overline{x}| + \cdots + |x_n - \overline{x}|}{n} = \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}| \qquad (4.10)$$

Para o cálculo do desvio médio de um conjunto X de observações (dados) distribuídos em intervalos de classes, consideremos o ponto médio x_i, como na definição 4.1.10, e a frequência absoluta f_{ai} do i-ésimo intervalo. Assim, determinamos o desvio médio por:

$$\mathrm{dm(X)} = \frac{f_{a1}.|x_1 - \overline{x}\,| + f_{a2}.|x_2 - \overline{x}\,| + f_{a3}.|x_3 - \overline{x}\,| + \cdots + f_{an}.|x_n - \overline{x}\,|}{f_{a1} + f_{a2} + f_{a3} + \cdots + f_{an}} = \frac{1}{\sum_{i=1}^{n} f_{ai}} \sum_{i=1}^{n} f_{ai}.\,|x_1 - \overline{x}\,| \qquad (4.11)$$

4.1.28 Variância

A variância de um conjunto X de observações (dados), é a média dos quadrados dos desvios.

$$\mathrm{var(X)} = \frac{(x_1 - \overline{x})^2 + (x_2 - \overline{x})^2 + \cdots + (x_n - \overline{x})^2}{n} = \frac{1}{n}\sum_{i=1}^{n} (x_i - \overline{x})^2 \qquad (4.12)$$

Para o cálculo da variância de um conjunto X de observações (dados) distribuídos em intervalos de classes, consideremos o ponto médio x_i, como na definição 4.1.10, e a frequência absoluta f_{ai} do i-ésimo intervalo. Assim, determinamos a variância por:

$$\mathrm{var(X)} = \frac{f_{a1}.(x_1 - \overline{x})^2 + f_{a2}.(x_2 - \overline{x})^2 + f_{a3}.(x_3 - \overline{x})^2 + \cdots + f_{an}.(x_n - \overline{x})^2}{f_{a1} + f_{a2} + f_{a3} + \cdots + f_{an}} = \frac{1}{\sum_{i=1}^{n} f_{ai}} \sum_{i=1}^{n} f_{ai}.\,(x_i - \overline{x})^2 \qquad (4.13)$$

4.1.29 Desvio padrão

O desvio padrão é a raiz quadrada da variância.

$$\mathrm{dp(X)} = \sqrt{\mathrm{var(X)}} \qquad (4.14)$$

Calcula-se a variância, o desvio médio ou desvio padrão de um conjunto de dados observados, com o objetivo de indicar em média, o erro cometido ao tentar substituir o valor de cada observação x_i pelo valor da medida resumo $\overline{x}$ (média), do conjunto de dados.

É importante salientar que o desvio padrão deve ser preterível à variância para se estudar o desvio dos dados em relação à média pela vantagem de permitir uma medida interpretativa que é expressa na mesma unidade de medida da variável (kg, km, l, atm etc).

Exemplo: Novamente observam-se os valores cotados do dólar no primeiro trimestre de 2020 pelo polígono de frequência, Gráfico 8, onde já se conhece o valor médio R$ 4,485 da cotação, e vamos calcular o desvio médio, a variância e o desvio padrão desta distribuição.

I) para o desvio médio, temos:

$$\mathrm{dm}(X) = \frac{11.\,|4{,}095 - 4{,}485| + 17.\,|4{,}245 - 4{,}485| + 10.\,|4{,}395 - 4{,}485| + \cdots + 8.\,|5{,}145 - 4{,}485|}{11 + 17 + 10 + \cdots + 8}$$

$$\mathrm{dm}(X) = \frac{4{,}29 + 4{,}08 + 0{,}90 + \cdots + 5{,}28}{62}$$

$$\mathrm{dm}(X) = \frac{18{,}51}{62} = 0{,}30$$

II) para a variância, temos:

$$\mathrm{var}(X) = \frac{11.\,(4{,}095 - 4{,}485)^2 + 17.\,(4{,}245 - 4{,}485)^2 + 10.\,(4{,}395 - 4{,}485)^2 + \cdots + 8.\,(5{,}145 - 4{,}485)^2}{11 + 17 + 10 + \cdots + 8}$$

$$\mathrm{var}(X) = \frac{1{,}673 + 0{,}979 + 0{,}081 + \cdots + 3{,}485}{62}$$

$$\mathrm{var}(X) = \frac{7{,}671}{62} = 0{,}124$$

III) para o desvio padrão, temos:

$$\mathrm{dp}(X) = \sqrt{\mathrm{var}(X)} = \sqrt{0{,}124} = 0{,}35$$

Portanto, se substituirmos cada um dos 62 valores da cotação do dólar pelo seu valor médio de R$ 4,485, cometeremos um erro médio de R$ 0,30 pelo desvio médio e de R$ 0,35 pelo desvio padrão.

4.2 Teoria da Probabilidade

A Probabilidade é um ramo de estudo da matemática que quantifica a possibilidade ou chance de um acontecimento (evento) vir a se concretizar. É uma das ferramentas matemáticas que está à disposição das diversas Ciências que buscam sistematizar resultados observados para o embasamento de novas teorias. Segundo Triola (2008, p. 112) “A probabilidade é a base sobre a qual são construídos importantes métodos de inferência estatística”.

Aqui, chama-se a atenção para o significado de inferir, cuja derivação vem do latim *infere* e significa deduzir, entender, compreender. A inferência estatística consiste em um conjunto de métodos que permitem fazer extrapolações das informações e conclusões obtidas a partir de uma amostra da população. Para Andrade et al. (2015, p. 24) “Os métodos de Inferência Estatística permitem (1) estimar as características desconhecidas de uma população [...] e (2) testar se determinadas hipóteses sobre essas características desconhecidas são plausíveis [...]”.

Frequentemente leigos questionam como os pesquisadores podem predizer aquilo que vai acontecer num tempo futuro. Com base em conceitos matemáticos, a inferência estatística faz generalizações sobre as características de uma população através da análise das informações contidas na amostra. Para Bussab e Morettin (2010, p. 01) “[...] a essência da Ciência é a observação e seu objetivo básico é a inferência, que pode ser dedutiva (na qual se argumenta das premissas às conclusões) ou indutiva (por meio da qual se vai do específico ao geral)”. Matematicamente, pela modelagem de certo fenômeno, se estabelece uma relação entre as premissas e a conclusão que orientam estudos e posteriores tomadas de decisões.

Diante do que se coloca, apresentaremos aqui algumas definições e resultados da Teoria da Probabilidade que serão necessários posteriormente para fazer o desenvolvimento de uma proposta de aplicação de uma modelagem adequada aos objetivos do Ensino Médio.

4.2.1 Experimento aleatório

O experimento aleatório é aquele que se apresenta como resultado do acaso e que não pode ser previsto com certeza. Assim, se repetido nas mesmas condições, produzirá diversos resultados.

São exemplos de experimentos aleatórios:

- lançar um dado com faces numeradas de 1 a 6 e observar o número da face superior;
- lançar duas moedas idênticas simultaneamente e observar as sequências de caras e coroas obtidas;
- em sorteios que usam bolinhas numeradas para gerar um apostador contemplado, observar as sequências dos números obtidos.

4.2.2 Espaço amostral

O espaço amostral é o conjunto, indicado por Ω, de todos os resultados possíveis para um experimento aleatório.

Como exemplo, tome duas moedas idênticas cujas faces sejam representadas por cara (C) e coroa (K). Jogue-as simultaneamente e observe os resultados. Em uma sequência de lançamentos, pode-se obter:

- C na primeira moeda e C na segunda;
- C na primeira moeda e K na segunda;
- K na primeira moeda e C na segunda;
- K na primeira moeda e K na segunda.

Observados todos os resultados possíveis, teremos o espaço amostral Ω do experimento aleatório realizado:

$$\Omega = \{(C,C); (C,K); (K,C); (K,K)\}.$$

Observa-se que este experimento aleatório realizado gerou um espaço amostral finito, pois temos um total de quatro resultados possíveis. Todavia, existem experimentos aleatórios que podem gerar espaços amostrais infinitos como se vê a seguir.

Exemplo: Tome dez moedas idênticas cujas faces são cara e coroa. Jogue-as simultaneamente e anote os resultados até que ocorra a face cara em todas as dez moedas. Identificando novamente a face cara por C e a face coroa por K, obtemos os possíveis resultados:

1º) lançamento simultâneo: C, K, K, K, C, C, K, C, C, C;
2º) lançamento simultâneo: K, C, K, K, C, C, K, C, K, C;
3º) lançamento simultâneo: C, C, K, K, K, K, K, C, K, C;
4º) lançamento simultâneo: C, C, C, C, C, C, K, C, C, C;
5º) lançamento simultâneo: C, K, K, K, C, C, K, C, C, C;

.
.
.

Observados alguns resultados possíveis, temos o espaço amostral Ω que está em formação, oriundo do experimento aleatório proposto:

Ω = {(C, K, K, K, C, C, K, C, C, C); (K, C, K, K, C, C, K, C, K, C); (C, C, K, K, K, K, K, C, K, C); (C, C, C, C, C, C, K, C, C, C); ...}.

Nota-se que apesar do resultado (C, C, C, C, C, C, C, C, C, C) ser possível de ocorrer em algum momento, também se observa que pela aleatoriedade do experimento, pode ser que ele nunca ocorra; gerando assim um espaço amostral infinito.

4.2.3 Evento

Chamamos de evento a todo subconjunto do espaço amostral Ω. E o designaremos por uma letra maiúscula do alfabeto: A, B, ..., Z.

Exemplo: Tome um dado com faces numeradas de 1 a 6. Jogue este dado e observe a face voltada para cima. Note que o espaço amostral deste resultado é Ω = {1; 2; 3; 4; 5; 6}.
Logo, alguns eventos podem ser:

- evento A: a face voltada para cima ser um número primo A = {2; 3; 5};
- evento Q: a face voltada para cima ser um número primo e par Q = {2};
- evento H: a face voltada para cima ser um número ímpar H = {1; 3; 5}.

4.2.4 Probabilidade

Considere um experimento aleatório que gere “n” resultados distintos constituindo um espaço amostral. Agora, tome um subconjunto A do conjunto Ω (A ⊂ Ω). Seja “s” o número de elementos do conjunto A.

Definimos a probabilidade **P** de ocorrer o evento **A** (que indicamos por P(A)), como o valor da razão entre “**s**” e “**n**”:

$$\mathrm{P(A)} = \frac{\mathrm{s}}{\mathrm{n}} \tag{4.15}$$

Exemplo: Tome duas moedas idênticas cujas faces são cara (C) e coroa (K). Ao jogar estas moedas sobre uma mesa, qual é a probabilidade de se obter cara na face superior em ambas?
Observados todos os resultados possíveis, temos o espaço amostral Ω do experimento aleatório: Ω = {(C,C); (C,K); (K,C); (K,K)}. E tomemos A = {(C,C)}.
Assim, temos s = 1 e n = 4 que nos fornece: $\mathrm{P(A)} = \frac{\mathrm{s}}{\mathrm{n}} = \frac{1}{4} = 0{,}25$. Logo, a probabilidade estimada é de 25% de chance para a ocorrência desejada.

Da definição de probabilidade que se apresenta, observamos algumas propriedades:

- tomado que A ⊂ Ω, temos $s \leq n$ onde $0 < \mathrm{P(A)} \leq 1$;
- caso A ⊄ Ω, falamos que A é um evento impossível e temos que $\mathrm{A} \cap \Omega = \emptyset$ ($\mathrm{s} = 0$ e $\mathrm{n} \neq 0$), onde $\mathrm{P(A)} = \frac{\mathrm{s}}{\mathrm{n}} = \frac{0}{\mathrm{n}} = 0$;
- caso A = Ω (s = n), falamos que A é um evento certo e temos que $\mathrm{P(A)} = \frac{\mathrm{s}}{\mathrm{n}} = \frac{\mathrm{n}}{\mathrm{n}} = 1 = \mathrm{P}(\Omega)$;

4.2.5 Adição das probabilidades (teorema)

Sejam A e B eventos mutuamente excludentes (isto é, $A \cap B = \emptyset$) tal que $A \cup B = \Omega$, então $P(A \cup B) = P(A) + P(B)$.

Demonstração:

Por hipótese, $A \cap B = \emptyset$ e $A \cup B = \Omega$.

Considere a cardinalidade dos eventos e do espaço amostral Ω como n(A), n(B) e n(Ω).

Logo temos que $n(\Omega) = n(A \cup B) = n(A) + n(B)$

$$n(A \cup B) = n(A) + n(B) \Leftrightarrow \frac{n(A \cup B)}{n(\Omega)} = \frac{n(A)}{n(\Omega)} + \frac{n(B)}{n(\Omega)} \Leftrightarrow P(A \cup B) = P(A) + P(B)$$

como queríamos demonstrar

Como consequência da demonstração acima, temos que o complementar de um evento A em relação ao espaço amostral Ω, representado por $\overline{A}$, é o evento em que todos os seus resultados são distintos dos resultados do evento A, ou seja, $A \cap \overline{A} = \emptyset$ e $A \cup \overline{A} = \Omega$; assim:

$$1 = P(\Omega) = P(A \cup \overline{A}) = P(A) + P(\overline{A}) \text{ onde } P(\overline{A}) = 1 - P(A).$$

4.2.6 União das probabilidades (teorema)

Sejam A e B eventos de um espaço amostral Ω tal que $A \cap B \neq \emptyset$, então $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Demonstração:

Por hipótese, $A \cap B \neq \emptyset$. Assim: $n(A \cup B) = n(A) + n(B) - n(A \cap B)$ onde $n(A \cup B)$, $n(A)$, $n(B)$ e $n(A \cap B)$ representam a cardinalidade dos eventos e n(Ω) a cardinalidade do espaço amostral. Daí:

$$n(A \cup B) = n(A) + n(B) - n(A \cap B)$$

$$\frac{n(A \cup B)}{n(\Omega)} = \frac{n(A)}{n(\Omega)} + \frac{n(B)}{n(\Omega)} - \frac{n(A \cap B)}{n(\Omega)}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

como queríamos demonstrar

Analogamente, podemos mostrar que para A, B e C eventos de um espaço amostral Ω tais que $A \cap B \neq \emptyset$, $A \cap C \neq \emptyset$, $B \cap C \neq \emptyset$ e $A \cap B \cap C \neq \emptyset$, então:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

4.2.7 Probabilidade condicional

Tome um conjunto Ω como o espaço amostral de um experimento aleatório e considere dois subconjuntos (eventos) A e B de Ω, tais que $A \cap B \neq \emptyset$. Indicamos como P(A|B) a probabilidade do evento A ocorrer, uma vez que o evento B tenha ocorrido. Isto é, calculamos P(A|B) como se o evento B fosse o novo espaço amostral dentro do qual queremos calcular a probabilidade do evento A.

Assim, definimos:

$$\mathrm{P(A|B)} = \frac{\mathrm{n(A{\cap}B)}}{\mathrm{n(B)}} = \frac{\frac{\mathrm{n(A{\cap}B)}}{n(\Omega)}}{\frac{\mathrm{n(B)}}{n(\Omega)}} = \frac{\mathrm{P(A{\cap}B)}}{\mathrm{P(B)}} \text{ com } \mathrm{P(B)} > 0$$

Observa-se que: $\mathrm{P(A|B)} = \frac{\mathrm{P(A{\cap}B)}}{\mathrm{P(B)}} \Leftrightarrow \mathrm{P(A \cap B)} = \mathrm{P(A|B).P(B)}$. Logo, a probabilidade da ocorrência simultânea de dois eventos A e B que indicamos por P(A∩B), é igual ao produto da probabilidade de um deles pela probabilidade do outro, dada a probabilidade do segundo $\mathrm{P(B)} > 0$.

Para três eventos A, B e C de um espaço amostral Ω, tal que $A \cap B \cap C \neq \emptyset$, temos:

$$\mathrm{P(A \cap B \cap C)} = \mathrm{P(A|B).P(B).P(C|A \cap B)}$$

Generalizando para uma quantidade finita de eventos $A_1, A_2, A_3, A_4, \cdots, A_n$ subconjuntos de Ω ($A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n \neq \emptyset$), temos:

$$\mathrm{P}(\cap_{i=1}^{n} A_i) = \mathrm{P}(A_1)\,.\,\mathrm{P}(A_2|A_1)\,.\,\mathrm{P}(A_3|A_1 \cap A_2)\,.\cdots\,.\,\mathrm{P}(A_n|A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_{n-1})$$

4.2.8 Eventos independentes

Considere o espaço amostral (conjunto Ω) de um experimento aleatório e tome dois eventos, A e B, pertencentes a ele. Pela probabilidade condicional, indicada por P(A|B), o conhecimento de que o evento B ocorreu, muda as chances do evento A ocorrer. Assim, nos casos especiais em que a ocorrência do evento B não intervir nas chances do evento A ocorrer, julga-se que os eventos A e B são independentes, onde: $\mathrm{P(A|B)} = \mathrm{P(A)}$.

Colocando-se de outra forma, dois eventos A e B são independentes quando:

$$\mathrm{P(A \cap B)} = \mathrm{P(A|B).P(B)} = \mathrm{P(A)\,.\,P(B)}$$

4.2.9 Produto das probabilidades (teorema)

Se A e B são eventos independentes de um espaço amostral Ω, então $P(A \cap B) = P(A) \, . \, P(B)$.

Demonstração:

Como A e B são eventos independentes, temos que P(A|B)=P(A) pois o evento A não é afetado pelo evento B. Logo: $P(A \cap B) = P(A|B) . P(B) = P(A) . P(B)$.

como queríamos demonstrar

Generalizando, sejam $A_1, A_2, A_3, A_4, \dots, A_n$ eventos independentes de um espaço amostral Ω, então $P(A_1 \cap A_2 \cap A_3 \cap \dots \cap A_n) = P(A_1) \, . \, P(A_2) \, . \, P(A_3) \, . \, \dots \, . \, P(A_n)$.

4.2.10 Teorema de Bayes

Sejam A e B eventos de um espaço amostral Ω, tal que: $A \neq \emptyset$, $B \neq \emptyset$, $A \cap B \neq \emptyset$, $A \neq B$ e $A \cup B = \Omega$. Então, temos que

$$P(A) = P(A|B) . P(B) + P(A|B^c) . P(B^c) \quad \text{ou} \quad P(B) = P(B|A) . P(A) + P(B|A^c) . P(A^c)$$

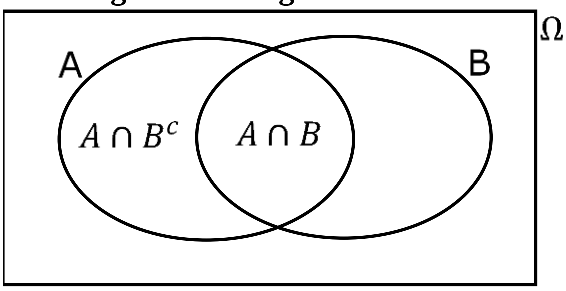
onde A^c é o complementar de A em relação a Ω e B^c é o complementar de B em relação a Ω.

Vamos demonstrar $P(A) = P(A|B) . P(B) + P(A|B^c) . P(B^c)$, pois a outra será de forma análoga.

Demonstração:

Por hipótese: $A \neq \emptyset$, $B \neq \emptyset$, $A \cap B \neq \emptyset$, $A \neq B$ e $A \cup B = \Omega$

Figura 5 – Diagrama de Venn



Fonte: Elaborado pelo autor (2020)

i) como $A \cap B$ e $A \cap B^c$ são eventos mutuamente excludentes, temos pela adição das probabilidades que $P(A) = P(A \cap B) + P(A \cap B^c)$

ii) $P(A|B) = \frac{P(A \cap B)}{P(B)} \Leftrightarrow P(A \cap B) = P(A|B) . P(B)$

iii) $P(A|B^c) = \frac{P(A \cap B^c)}{P(B^c)} \Leftrightarrow P(A \cap B^c) = P(A|B^c).P(B^c)$

Substituindo ii e iii em i temos:

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B).P(B) + P(A|B^c).P(B^c)$$

como queríamos demonstrar

4.2.11 Distribuições de probabilidade para uma variável aleatória discreta

Uma distribuição de probabilidade para uma variável aleatória discreta x, associa cada valor que ela pode assumir com sua respectiva probabilidade $P(x)$. Segundo Triola (2008, p. 161) “Uma variável aleatória discreta tem ou um número finito de valores ou uma quantidade enumerável de valores, onde enumerável se refere ao fato de que podem existir infinitos valores, mas que podem ser associados a um processo de contagem”.

A distribuição de probabilidade $P(x_i)$, em que a variável aleatória discreta x_i assume todos os valores possíveis quando $i = 1, 2, \dots, n$; tem que satisfazer as seguintes propriedades:

- $0 \leq P(x_i) \leq 1$;
- $\sum_{i=1}^{n} P(x_i) = 1$.

4.2.11.1 Distribuição de probabilidade binomial

De acordo com Triola (2008), a distribuição de probabilidade binomial nos permite o cálculo da probabilidade para a circunstância em que se tenham duas categorias bem definidas e independentes: o sucesso e o fracasso.

A probabilidade $P(x_i)$ de se obter x_i sucessos em n tentativas, é dada por:

$$P(x_i) = \frac{n!}{(n-x_i)!\,x_i!} \ .\ p^{x_i}\ .\ (1-p)^{n-x_i} \quad \text{para } x_i = 0, 1, 2, \dots, n \ \text{ e } \ i = 0, 1, 2, \dots, n \qquad (4.16)$$

onde: **p** é a probabilidade do sucesso e **(1-p)** é a probabilidade do fracasso.

5 ANÁLISE ESTATÍSTICA DE UM CONJUNTO DE DADOS

Nesse capítulo faremos análises de bancos de dados reais e apresentaremos discussões embasadas pelos conceitos da Estatística e da Probabilidade apresentados ao longo da dissertação. Por meio de investigações e observações, busca-se a compreensão e uma posterior descrição do conjunto de dados agrupados. Ressalta-se que o primeiro passo é definir o tema de interesse que se torna o objeto de pesquisa. O procedimento estatístico que será usado, dependerá da natureza deste tema e das informações presentes no banco de dados. Segundo Bussab e Morettin (2010, p. 01)

> Em alguma fase de seu trabalho, o pesquisador depara-se com o problema de analisar e entender um conjunto de dados relevante ao seu particular objeto de estudos. Ele necessitará trabalhar os dados para transformá-los em informações, para compará-los com outros resultados, ou ainda para julgar sua adequação a alguma teoria.

Ao final de um trabalho, espera-se que os dados possam trazer novas informações agregando conhecimentos que apontem para futuras linhas de pesquisa e ideias que se transformem em projetos e ações contribuindo positivamente no desenvolvimento das ações humanas.

Neste sentido desenvolve-se aqui, uma breve análise de três bancos de dados com informações pertinentes ao período de 2009 a 2019 para verificar, em cada estudo, das realidades de que se tratam. O primeiro banco de dados trabalhado é sobre o recenseamento demográfico da população brasileira e foi extraído do banco de tabelas estatísticas Sidra, do site Instituto Brasileiro de Geografia e Estatística (IBGE)[6]. Os outros bancos tratam, um do número de imunizações contra o Sarampo, a Caxumba e a Rubéola, e o outro do número de casos registrados de infecção por Sarampo. Todos com dados amostrais de abrangência nacional e extraídos do site do Datasus (2020)[7]. A exploração e classificação das informações contidas nesses bancos foram construídas no *software* Excel utilizando diferentes tipos de gráficos e tabelas.

---

[6] https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas
[7] http://www2.datasus.gov.br/DATASUS/index.php?area=02

5.1 Banco de Dados: População Brasileira

De acordo com informações do site do IBGE[8], o censo demográfico é uma fonte para se ter referência das condições de vida da população do país, tendo como base a pessoa residente no território nacional, na data da coleta de dados. O primeiro censo da população do Brasil foi realizado em 1808 para atender interesses militares. Em 1940 houve a reestruturação dos serviços de estatística do País com a criação do Conselho Nacional de Estatística (CNE) e o Conselho Nacional de Geografia (CNG), que integram o IBGE.

Segundo o IBGE[3], os censos são realizados a cada 10 anos, através da aplicação de um Questionário Básico que colhe dados sobre o domicílio, os moradores e características gerais da população como as informações sobre a estrutura etária, sexo, cor ou raça, distribuição espacial da população, alfabetização, registro de nascimento, domicílios ocupados ou domicílios vagos, propriedade do imóvel, saneamento ambiental, energia elétrica e renda média per capita.

Já a estimativa da população, conforme informações do site do IBGE[9], fornece valores aproximados para a população dos municípios e estados do País, a partir dos censos demográficos e é divulgada no dia 1º de julho do ano corrente. Esta estimativa é fornecida desde 1975 e a partir de 1992, publicada no Diário Oficial da União. Seu cálculo é realizado por um método matemático criado por João Lira Madeira e Celso Cardoso da Silva Simões em 1972, chamado AiBi, em que

> (...) utiliza como insumos básicos as populações obtidas das Projeções da População para o Brasil e as Unidades da Federação mais recentes, bem como o crescimento populacional de cada Município na última década, delineado pelas respectivas populações recenseadas nos dois últimos Censos Demográficos realizados (IBGE)[4].

Pensando na importância das informações citadas acima, iniciamos nossa análise explorando a distribuição da população brasileira nos âmbitos estaduais e regionais no período de 2009 a 2019. Os dados coletados, são expostos, inicialmente, em uma tabela para que o leitor possa ter informações quantitativas e gerais acerca da distribuição populacional. Por mais simples e resumido que seja a apresentação dos dados agrupados nesta tabela, é clara a necessidade de se resumir o conjunto de observações em gráficos e tabelas com informações

[8] https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html?=&t=o-que-e
[9] https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=o-que-ec

mais diretas. O objetivo é identificar as relações existentes entre os dados para que se possa, em um segundo momento, analisar as informações a partir de alguma teoria.

Considerando onze tabelas com dados absolutos da população brasileira residente no território nacional extraídas do site do IBGE – SIDRA[10], organizou-se a tabela 7 que agrupa as seguintes informações:

- valor absoluto da população por ano e o total populacional nas colunas;
- por região, unidades federativas e o distrito federal nas linhas.

[10] https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas

Tabela 7 – Distribuição da população brasileira no período de 2009 a 2019

DISTRIBUIÇÃO DA POPULAÇÃO BRASILEIRA POR REGIÃO, UNIDADE FEDERATIVA E O DISTRITO FEDERAL

ANO		2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019
REGIÃO NORTE	AC	691.132	732.793	746.386	758.786	776.463	790.101	803.513	816.687	829.619	869.265	881.935
	AP	626.609	668.689	684.309	698.602	734.996	750.912	766.679	782.295	797.722	829.494	845.731
	AM	3.393.369	3.480.937	3.538.387	3.590.985	3.807.921	3.873.743	3.938.336	4.001.667	4.063.614	4.080.611	4.144.597
	PA	7.457.119	7.588.078	7.688.593	7.822.205	7.999.729	8.104.880	8.206.923	8.305.359	8.366.628	8.513.497	8.602.865
	RO	1.503.928	1.560.501	1.576.455	1.590.011	1.728.214	1.748.531	1.768.204	1.787.279	1.805.788	1.757.589	1.777.225
	RR	421.499	451.227	460.165	469.524	488.072	496.936	505.665	514.229	522.636	576.568	605.761
	TO	1.292.051	1.383.453	1.400.892	1.417.694	1.478.164	1.496.880	1.515.126	1.532.902	1.550.194	1.555.229	1.572.866
	Total	**15.385.707**	**15.865.678**	**16.095.187**	**16.347.807**	**17.013.559**	**17.261.983**	**17.504.446**	**17.740.418**	**17.936.201**	**18.182.253**	**18.430.980**
REGIÃO NORDESTE	AL	3.156.108	3.120.922	3.143.384	3.165.472	3.300.935	3.321.730	3.340.932	3.358.963	3.375.823	3.322.820	3.337.357
	BA	14.637.364	14.021.432	14.097.534	14.175.341	15.044.137	15.126.371	15.203.934	15.276.566	15.344.447	14.812.617	14.873.064
	CE	8.547.809	8.448.055	8.530.155	8.606.005	8.778.576	8.842.791	8.904.459	8.963.663	9.020.460	9.075.649	9.132.078
	MA	6.367.138	6.569.683	6.645.761	6.714.314	6.794.301	6.850.884	6.904.241	6.954.036	7.000.229	7.035.055	7.075.181
	PB	3.769.977	3.766.834	3.791.315	3.815.171	3.914.421	3.943.885	3.972.202	3.999.415	4.025.558	3.996.496	4.018.127
	PE	8.810.256	8.796.032	8.864.906	8.931.028	9.208.550	9.277.727	9.345.173	9.410.336	9.473.266	9.496.294	9.557.071
	PI	3.145.325	3.119.015	3.140.328	3.160.748	3.184.166	3.194.718	3.204.028	3.212.180	3.219.257	3.264.531	3.273.227
	RN	3.137.541	3.168.133	3.198.657	3.228.198	3.373.959	3.408.510	3.442.175	3.474.998	3.507.003	3.479.010	3.506.853
	SE	2.019.679	2.068.031	2.089.819	2.110.867	2.195.662	2.219.574	2.242.937	2.265.779	2.288.116	2.278.308	2.298.696
	Total	**53.591.197**	**53.078.137**	**53.501.859**	**53.907.144**	**55.794.707**	**56.186.190**	**56.560.081**	**56.915.936**	**57.254.159**	**56.760.780**	**57.071.654**
REGIÃO CENTRO-OESTE	DF	2.606.885	2.562.963	2.609.998	2.648.532	2.789.761	2.852.372	2.914.830	2.977.216	3.039.444	2.974.703	3.015.268
	GO	5.926.300	6.004.045	6.080.716	6.154.996	6.434.048	6.523.222	6.610.681	6.695.855	6.778.772	6.921.161	7.018.354
	MT	3.001.692	3.033.991	3.075.936	3.115.336	3.182.113	3.224.357	3.265.486	3.305.531	3.344.544	3.441.998	3.484.466
	MS	2.360.498	2.449.341	2.477.542	2.505.088	2.587.269	2.619.657	2.651.235	2.682.386	2.713.147	2.748.023	2.778.986
	Total	**13.895.375**	**14.050.340**	**14.244.192**	**14.423.952**	**14.993.191**	**15.219.608**	**15.442.232**	**15.660.988**	**15.875.907**	**16.085.885**	**16.297.074**
REGIÃO SUDESTE	ES	3.487.199	3.512.672	3.547.055	3.578.067	3.839.366	3.885.049	3.929.911	3.973.697	4.016.356	3.972.388	4.018.650
	MG	20.033.665	19.595.309	19.728.701	19.855.332	20.593.356	20.734.097	20.869.101	20.997.560	21.119.536	21.040.662	21.168.791
	RJ	16.010.429	15.993.583	16.112.678	16.231.365	16.369.179	16.461.173	16.550.024	16.635.996	16.718.956	17.159.960	17.264.943
	SP	41.384.039	41.252.160	41.587.182	41.901.219	43.663.669	44.035.304	44.396.484	44.749.699	45.094.866	45.538.936	45.919.049
	Total	**80.915.332**	**80.353.724**	**80.975.616**	**81.565.983**	**84.465.570**	**85.115.623**	**85.745.520**	**86.356.952**	**86.949.714**	**87.711.946**	**88.371.433**
REGIÃO SUL	PR	10.686.247	10.439.601	10.512.349	10.577.755	10.997.465	11.081.692	11.163.018	11.242.720	11.320.892	11.348.937	11.433.957
	RS	10.914.128	10.695.532	10.733.030	10.770.603	11.164.043	11.207.274	11.247.972	11.286.500	11.322.895	11.329.605	11.377.239
	SC	6.118.743	6.249.682	6.317.054	6.383.286	6.634.254	6.727.148	6.819.190	6.910.553	7.001.161	7.075.494	7.164.788
	Total	**27.719.118**	**27.384.815**	**27.562.433**	**27.731.644**	**28.795.762**	**29.016.114**	**29.230.180**	**29.439.773**	**29.644.948**	**29.754.036**	**29.975.984**
TOTAL POR ANO		**191.506.729**	**190.732.694**	**192.379.287**	**193.976.530**	**201.062.789**	**202.799.518**	**204.482.459**	**206.114.067**	**207.660.929**	**208.494.900**	**210.147.125**

Fonte: Adaptado https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas
Acesso em: 10 de setembro de 2020.

Dados do censo demográfico de 2019, registra uma população absoluta de 210.147.125 habitantes. Analisando a tabela 7, percebe-se que a região sudeste é a mais populosa com 88.371.433 habitantes, cerca de 42% da população brasileira. O estado de São Paulo tem 45.919.049 habitantes, o maior índice populacional entre os 26 estados e o Distrito Federal. Em contra partida, o estado de Roraima tem o menor índice absoluto com 605.761 habitantes. Para se ter uma representatividade destes números populacionais observados, em relação aos

210.147.125 habitantes, São Paulo possui 21,85% desta população nacional, enquanto Roraima, 0,29%. A tabela 8 mostra a distribuição da população brasileira, por região, em 2019.

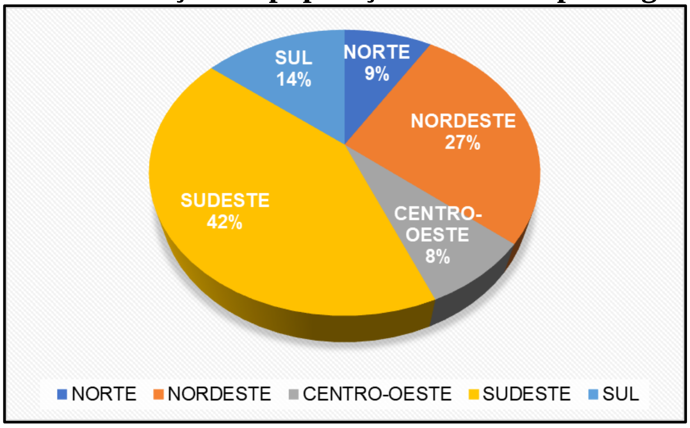
Tabela 8 – Valor absoluto da população por região em 2019

NORTE	18.430.980
NORDESTE	57.071.654
CENTRO-OESTE	16.297.074
SUDESTE	88.371.433
SUL	29.975.984
TOTAL	210.147.125

Fonte: Adaptado https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas
Acesso em: 10 de setembro de 2020.

Nota-se que as regiões nordeste e sudeste são as mais populosas, enquanto a região centro-oeste tem 16.297.074 habitantes, o menor índice populacional absoluto. Apesar desta tabela ser informativa, não temos uma visão da representatividade desta distribuição populacional. O Gráfico 11 mostra, de forma proporcional, a distribuição da população pelas cinco regiões para que se tenha uma percepção da subdivisão populacional pelo território brasileiro.

Gráfico 11 – Distribuição da população brasileira por região em 2019



Fonte: Adaptado https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas
Acesso em: 10 de setembro de 2020.

Pela análise do Gráfico11 em conjunto com a tabela 7, percebe-se que as regiões sudeste e nordeste tem os maiores índices populacionais em todo o período. Aqui, apontamos a necessidade de uma investigação para entender as possíveis causas desta distribuição.

No período de 11 anos, o qual se trata a tabela 7, é possível observar que a região sudeste teve o maior crescimento populacional com um aumento de 7.456.101 habitantes. Em termos absolutos, a região sul teve o menor crescimento populacional com 2.256.866 habitantes. Já o Norte teve o maior crescimento percentual populacional (19,79%) e a região nordeste (6,49%), o menor. Estas informações podem ser melhor visualizadas na tabela 9, que apresenta um resumo dos dados referentes ao crescimento percentual da população nas cinco regiões.

Tabela 9 – Análise do crescimento populacional por região

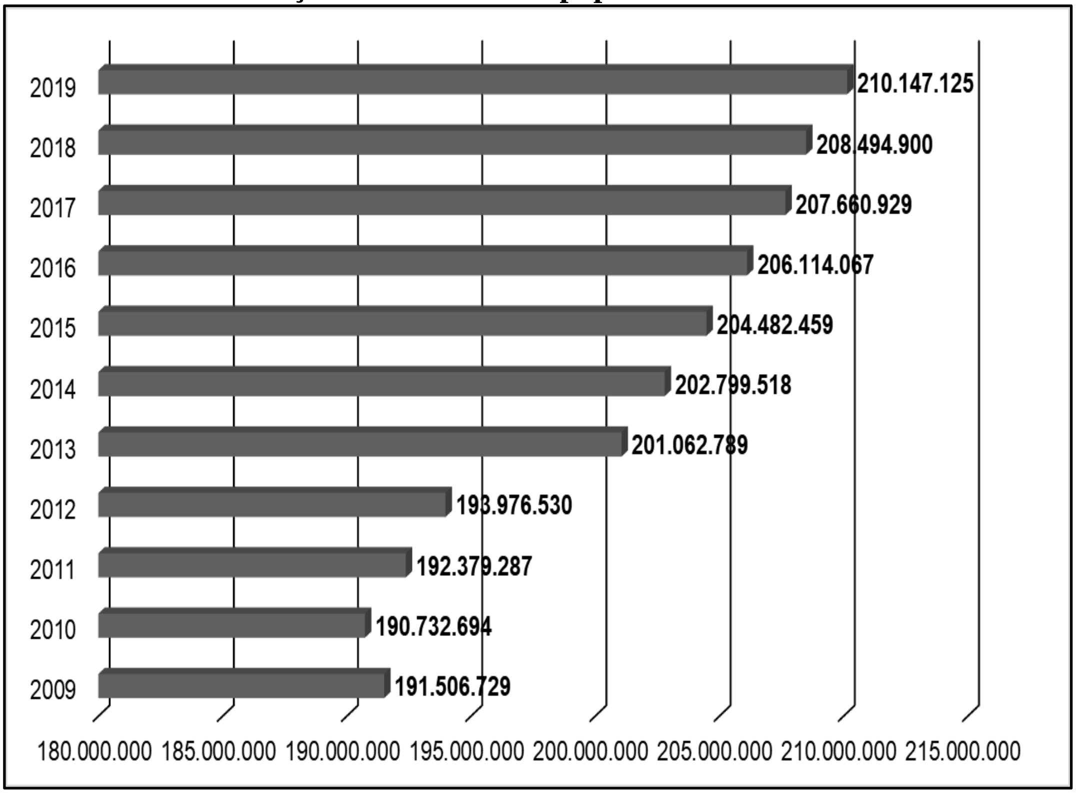
CRESCIMENTO POPULACIONAL NO PERÍODO				
REGIÃO	POPULAÇÃO EM 2009	POPULAÇÃO EM 2019	CRESCIMENTO ABSOLUTO	CRESCIMENTO PERCENTUAL
Norte	15.385.707	18.430.980	3.045.273	19,79%
Nordeste	53.591.197	57.071.654	3.480.457	6,49%
Centro-Oeste	13.895.375	16.297.074	2.401.699	17,28%
Sudeste	80.915.332	88.371.433	7.456.101	9,21%
Sul	27.719.118	29.975.984	2.256.866	8,14%

Fonte: Adaptado https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas
Acesso em: 10 de setembro de 2020.

Em média, o crescimento por região foi de 3.728.079 habitantes. Ressalta-se que as regiões norte e centro-oeste têm as maiores taxas percentuais de crescimento que se justificam pelos menores índices populacionais absolutos de cada uma destas regiões. Assim, pequenas variações em números absolutos populacionais acabam por gerar índices percentuais de crescimento que chamam a atenção, mas que não refletem a realidade local dando a impressão de uma explosão demográfica.

A tabela 7 sobre a Distribuição da População Brasileira por Região e Unidade Federativa, é densa de informações quantitativas. Para facilitar a análise dos dados sobre o crescimento demográfico no período, o Gráfico 12 apresenta os valores absolutos da população por ano.

Gráfico 12 – Distribuição do crescimento populacional brasileiro de 2009 a 2019



Fonte: Adaptado https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas
Acesso em: 10 de setembro de 2020.

Neste período de onze anos observados, o crescimento percentual da população é de aproximadamente 9,73%. Para entendê-lo, organiza-se a tabela 10 com dez classes sobre o crescimento populacional por biênio, cuja amplitude da amostra é de 18.640.396 habitantes. Esta tabela permite uma análise estatística dos dados em termos de crescimento absoluto do número de habitantes no Brasil.

Tabela 10 – Crescimento absoluto da população por biênio

2009-2010	-774.035
2010-2011	1.646.593
2011-2012	1.597.243
2012-2013	7.086.259
2013-2014	1.736.729
2014-2015	1.682.941
2015-2016	1.631.608
2016-2017	1.546.862
2017-2018	833.971
2018-2019	1.652.225

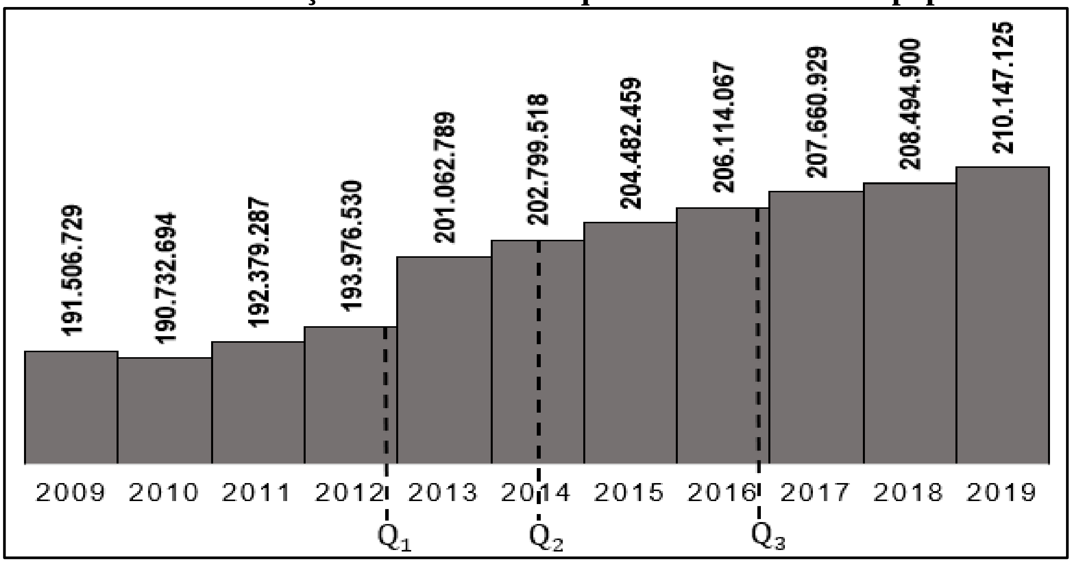
Fonte: Adaptado https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas
Acesso em: 10 de setembro de 2020.

Nota-se que ocorreu uma diminuição populacional no biênio 2009-2010 em 774.035 habitantes. Percentualmente é uma diminuição de aproximadamente 0,4% na comparação com a população de 2009. Aqui, apontamos a necessidade de uma investigação para compreender o que possa ter influenciado este fenômeno. Caso a pesquisa não aponte dados que justifiquem esta queda populacional, podemos considerá-la como *outlier* ou um valor atípico quando comparado ao crescimento dos anos subsequentes.

Por biênio, a população tem um crescimento médio de 1.864.039,60 habitantes e o desvio médio destes dados, que é o erro médio ao tentar substituir os valores da tabela pela sua média, foi de 1.044.443,88 habitantes. Chama a atenção um desvio médio tão alto, mas entende-se seu valor quando se analisam os biênios de 2009-2010 com uma retração no crescimento, e 2012-2013 cujo crescimento populacional foi de 7.086.259 habitantes, que destoa dos demais e onde apontamos novamente a necessidade de uma investigação para determinar a causa deste dado discrepante e inesperado, quando comparado aos demais no mesmo período.

Observa-se também pelo Gráfico 12 que o crescimento populacional brasileiro é assimétrico. Esta distribuição assimétrica já era esperada pois, tem-se um crescimento natural da população em praticamente todo o período. Assim, constrói-se o Gráfico 13 com barras verticais e adjacentes que favorecem uma análise visual deste crescimento populacional e de sua distribuição que é assimétrica à esquerda. Nota-se que outra vantagem deste gráfico é a determinação da mediana da distribuição com 202.799.518 habitantes.

Gráfico 13 – Distribuição assimétrica à esquerda do crescimento populacional



Fonte: Adaptado https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas
Acesso em: 10 de setembro de 2020.

Para não fazer esta classificação apenas visualmente, determinam-se os três quartis que dividem os dados amostrais do período em quatro partes contendo 25% da população em cada, e a seguir calcula-se o coeficiente de Bowley.

Usando o *software* Excel determinamos os quartis: $Q_1 = 197.153.124$, $Q_2 = 202.799.518$ e $Q_3 = 206.473.322$. Observa-se que o crescimento acima da média de 7.086.259 habitantes no biênio 2012-2013, está influenciando esta distribuição. Assim, temos que pelo menos 25% da distribuição da população no período está abaixo de Q_1 e os outros 75% estão acima. Nota-se que o primeiro quartil detém a maior parte do período observado com aproximadamente 36,4% dos 11 anos analisados, alongando a calda esquerda da distribuição.

No cálculo do coeficiente de Bowley, obtemos:

$$B = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{(Q_3 - Q_1)}$$

$$B = \frac{(206.473.322 - 202.799.518) - (202.799.518 - 197.153.124)}{(206.473.322 - 197.153.124)}$$

$$B = \frac{(3.673.804) - (5.646.394)}{(9.320.198)}$$

$$B = -0{,}212.$$

Como o coeficiente é negativo, sua classificação é de uma distribuição assimétrica da população à esquerda. Se o coeficiente estivesse mais próximo de zero, ter-se-ia uma distribuição aproximadamente simétrica sem o degrau desproporcional do biênio 2012-2013.

5.2 Banco de Dados: Sarampo

Segundo Stevanim (2018), em um artigo para Fundação Oswaldo Cruz (Fiocruz), o sarampo é uma doença infecciosa causada por um vírus da família *Paramyxoviridae* e gênero *Morbillivirus* cujos sintomas iniciais são: febre, tosse, irritação ocular e corrimento no nariz. Após a manifestação destes sintomas, é comum aparecerem manchas vermelhas pelo corpo, infecção nos ouvidos, ataques convulsivos, infecções no encéfalo, diarreia e pneumonia. As formas mais graves da doença se manifestam em recém-nascidos, gestantes e portadores de imunodeficiências. A transmissão se dá por gotículas suspensas no ar, quando um indivíduo infectado tosse, espirra, fala ou até mesmo respira.

Em 2016, a Organização Pan-Americana da Saúde concedeu ao Brasil o certificado de eliminação da circulação do vírus no território nacional. Isto não quer dizer que a doença deixou de existir e sim, que está com baixos níveis de contágio na população por uma transmissão não sustentada e poucos registros de morbidade hospitalar. A morbidade é um termo técnico-científico usado para indicar a quantidade de indivíduos afetados por uma doença em determinado local e momento (STEVANIM, 2018).

Uma transmissão é considerada sustentada quando se tem surtos da doença por um período maior que 12 meses. Em 2018 a doença retornou ao país com surtos no Amazonas e em Roraima. Na época, a falta de vacinação foi apontada como a principal causa para o surto, visto que a vacina é a principal forma de combate à doença (STEVANIM, 2018).

A prevenção desta doença se dá pela vacinação em massa da população. De acordo com Stevanim (2018) para o site da Fiocruz, é recomendado que a criança receba a primeira dose da vacina aos 12 meses de idade, com a tríplice viral (sarampo, caxumba e rubéola) e tenha um reforço aos 15 meses com a tetra viral (sarampo, caxumba, rubéola e varicela). Crianças, adolescentes e adultos que não foram vacinados no momento correto, devem receber esta imunização procurando um posto de saúde para atualizar seu cartão de vacina.

Dados extraídos do Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)[11], mostram os registros de infecção pelo vírus do sarampo no Brasil e o número de imunizações realizadas pela aplicação da vacina tríplice viral, no período de 2009 a 2019. A partir destas informações tem-se a tabela 11 com as notificações de internação hospitalar (morbidade hospitalar) por sarampo e a tabela 12 com doses aplicadas da vacina por região.

[11] http://www2.datasus.gov.br/DATASUS/index.php?area=02

Tabela 11 – Número de infecções pelo vírus do sarampo de 2009 a 2019

MORBIDADE HOSPITALAR - SARAMPO													
ANO		2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Total
REGIÃO NORTE	AC	0	0	0	0	0	0	0	0	0	0	0	**0**
	AP	0	0	0	0	1	1	0	0	0	0	0	**2**
	AM	0	0	0	0	0	0	0	1	0	733	4	**738**
	PA	0	2	3	2	0	1	0	2	0	14	20	**44**
	RO	0	0	1	0	2	1	0	0	0	5	3	**12**
	RR	1	0	0	0	0	0	0	0	0	68	1	**70**
	TO	1	0	0	0	0	0	1	1	0	0	1	**4**
	Total	**2**	**2**	**4**	**2**	**3**	**3**	**1**	**4**	**0**	**820**	**29**	
REGIÃO NORDESTE	AL	0	1	0	0	0	0	0	0	0	0	6	**7**
	BA	3	0	1	0	2	0	0	2	1	3	17	**29**
	CE	2	11	0	0	0	59	41	1	1	2	6	**123**
	MA	0	0	0	0	2	0	0	1	0	2	2	**7**
	PB	0	5	1	0	1	0	0	0	0	1	11	**19**
	PE	7	6	9	23	47	12	3	19	52	16	52	**246**
	PI	0	0	0	0	1	1	0	1	2	1	2	**8**
	RN	0	0	1	0	0	0	0	0	0	1	1	**3**
	SE	0	0	0	0	2	1	0	1	0	0	0	**4**
	Total	**12**	**23**	**12**	**23**	**55**	**73**	**44**	**25**	**56**	**26**	**97**	
REGIÃO CENTRO OESTE	DF	0	0	0	0	0	0	0	0	0	0	0	**0**
	GO	5	0	0	2	3	0	0	0	0	0	15	**25**
	MT	0	0	1	1	2	0	0	0	0	0	0	**4**
	MS	0	0	2	0	2	1	1	0	1	11	5	**23**
	Total	**5**	**0**	**3**	**3**	**7**	**1**	**1**	**0**	**1**	**11**	**20**	
REGIÃO SUDESTE	ES	0	3	1	1	2	0	1	0	0	2	9	**19**
	MG	1	2	2	9	1	2	0	1	0	22	66	**106**
	RJ	2	1	0	4	1	0	2	0	2	1	19	**32**
	SP	2	4	5	5	1	2	2	1	0	5	568	**595**
	Total	**5**	**10**	**8**	**19**	**5**	**4**	**5**	**2**	**2**	**30**	**662**	
REGIÃO SUL	PR	7	5	11	4	1	0	0	0	0	1	4	**33**
	RS	11	24	30	12	5	2	4	2	4	10	8	**112**
	SC	3	1	1	0	0	0	0	0	0	3	13	**21**
	Total	**21**	**30**	**42**	**16**	**6**	**2**	**4**	**2**	**4**	**14**	**25**	
Total por ano		**45**	**65**	**69**	**63**	**76**	**83**	**55**	**33**	**63**	**901**	**833**	
TOTAL DE CASOS REGISTROS NO PERÍODO:					**2.286**								

Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Tabela 12 – Imunizações contra o vírus do sarampo por região

DOSES DA TRÍPLICE VIRAL APLICADAS NO PERÍODO DE 2009 A 2019	
NORTE	10.891.978
NORDESTE	30.633.265
CENTRO-OESTE	7.560.859
SUDESTE	46.752.994
SUL	13.070.836

Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Diante dos registros da tabela 11, observa-se que houve surtos da doença nos estados do Amazonas com 733 casos no ano de 2018 e de São Paulo com 568 casos em 2019. Percebe-se a gravidade desta situação ao analisar todo o período com os 2.286 casos registrados desta morbidade hospitalar. Em termos probabilísticos, a chance de uma infecção pelo vírus do sarampo ter ocorrido no Amazonas em 2018 ou em São Paulo em 2019 é

$$\mathrm{P}(\mathrm{A} \cup \mathrm{B}) = \mathrm{P}(\mathrm{A}) + \mathrm{P}(\mathrm{B}) = \frac{733}{2.286} + \frac{568}{2.286} = \frac{1.301}{2.286} = 0{,}5691$$

onde A e B são os indivíduos infectados pelo sarampo no Amazonas (2018) ou em São Paulo (2019), respectivamente.

Neste período de 2009 a 2019 tem-se uma média de 208 casos de infecção por ano. Já em uma comparação por estado, a média de infecções foi de 85 casos.

Como já mencionado, a prevenção contra essa doença se dá pela vacinação em massa de uma população. A tabela 12 mostra a quantidade de dozes aplicadas da vacina tríplice viral nas cinco regiões em todo o período amostral. Nota-se que a maior parte das imunizações realizadas ocorreram na região sudeste com 46.752.994 doses aplicadas. Este dado já era esperado visto que se trata da região mais populosa do Brasil de acordo com os dados da tabela 8.

Em 2016, o Brasil recebeu o certificado de erradicação do Sarampo, por apresentar o menor número de casos registrados, 33 em todo o território nacional. Pela tabela 13, quinze das 27 unidades federativas (55,56%) não apresentaram registros da doença.

Tabela 13 – Frequência de casos por estado e o Distrito Federal em 2016

NÚMERO DE CASOS	UNIDADES FEDERATIVAS	PERCENTUAL
0	15	55,56%
1	8	29,63%
2	3	11,11%
19	1	3,70%

Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Ao analisar a tabela 11, nota-se que Pernambuco registrou 52 casos de sarampo em 2017 e 2019. Porém, a frequência de casos no período estudado, aponta que dos estados da região nordeste, este é o estado com o maior índice absoluto de morbidade hospitalar pelo sarampo, 246 registos. Isto equivale a 55,16% dos 446 infectados na região nordeste. Uma taxa de incidência tão alta, causa preocupação e aponta para a necessidade de uma investigação no intuito de determinar sua causa.

Pela tabela 14 nota-se que 74,07% das unidades federativas não tiveram caso de infecção por sarampo em 2017.

Tabela 14 – Frequência de casos por estado e o Distrito Federal em 2017

NÚMERO DE CASOS	UNIDADES FEDERATIVAS	PERCENTUAL
0	20	74,07%
1	3	11,11%
2	2	7,41%
4	1	3,70%
52	1	3,70%

Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Como mencionado anteriormente, em 2018 o Brasil perdeu o certificado de erradicação do Sarampo a nível nacional. Naquele ano, pela tabela 11, tem-se um surto com 901 casos registrados dos quais, 820 foram na região norte. Porém, pela tabela 15, em 20 das 27 unidades federativas, a contagem de casos registrados não passava de 5.

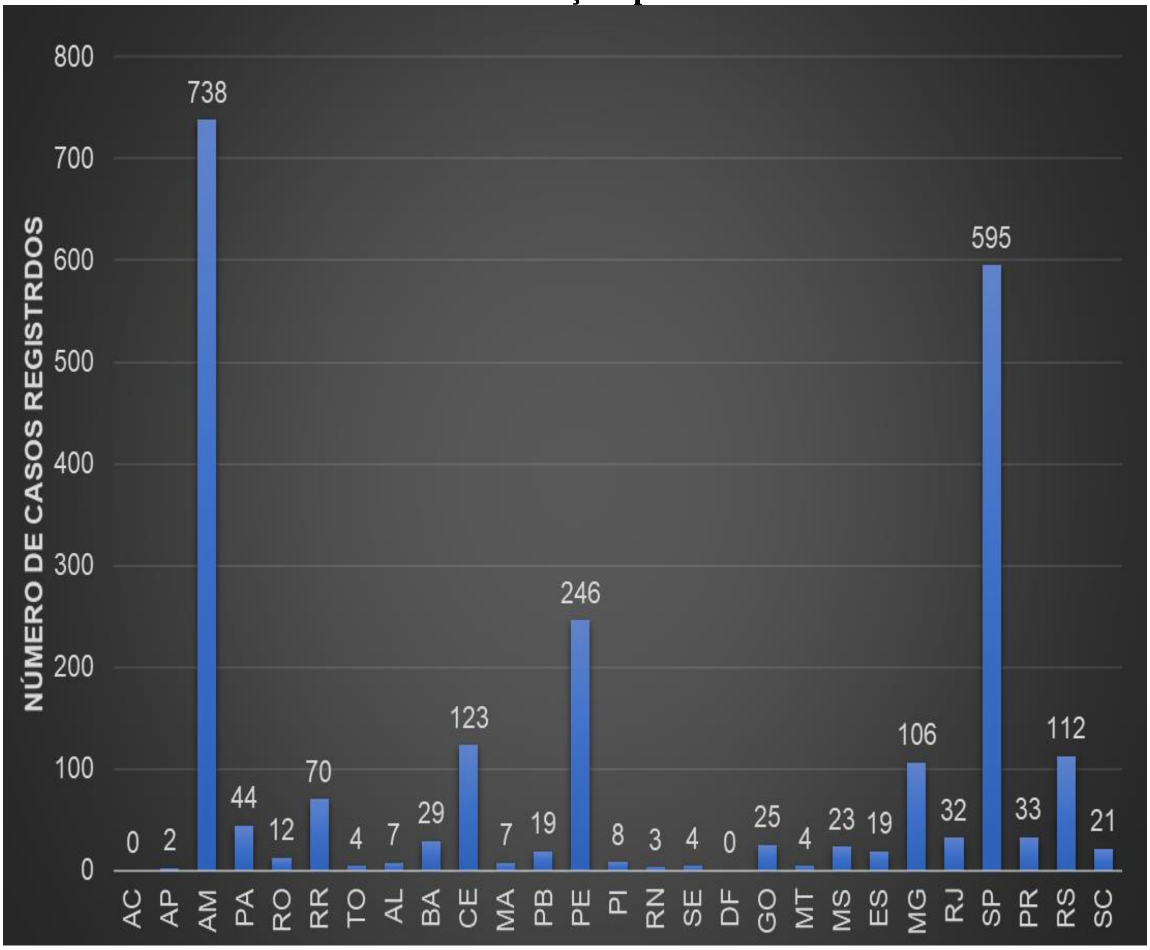
Tabela 15 – Frequência de casos por estado e o Distrito Federal em 2018

NÚMERO DE CASOS	UNIDADES FEDERATIVAS	PERCENTUAL
0	8	29,63%
1	5	18,52%
2	3	11,11%
3	2	7,41%
5	2	7,41%
10	1	3,70%
11	1	3,70%
14	1	3,70%
16	1	3,70%
22	1	3,70%
68	1	3,70%
733	1	3,70%

Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

O Gráfico 14 é construído a partir da tabela 11 onde destacam-se os estados do Amazonas e de São Paulo com uma quantidade assustadora de infecções no período, 738 e 595 respectivamente. Como já observado anteriormente, estes dados chamativos são influenciados pontualmente pelos surtos registrados em 2018 e 2019 com uma probabilidade de 56,91% de chance da morbidade ter ocorrido em um estado ou outro. Porém, ao se analisar individualmente estas localidades no período de 2016 a 2019, percebe-se que a chance de um indivíduo vir a se infectar pelo sarampo continua alta.

Gráfico 14 – Número de infecções por estado de 2009 a 2019



Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Pelos dados amostrais da tabela 11, a probabilidade de uma pessoa que foi infectada pelo vírus do sarampo residir no estado de São Paulo, sabendo que a infecção ocorreu no período de 2016 a 2019 é de:

$$P(A|B) = \frac{n(A \cap B)}{n(B)} = \frac{1 + 0 + 5 + 568}{33 + 63 + 901 + 833} = \frac{574}{1830} = 0{,}3137$$

onde A é o número de infectados pelo sarampo no estado de São Paulo nos 11 anos, e B é o número de infectados pelo sarampo a nível nacional no período de 2016 a 2019.

Já para no estado do Amazonas, esta probabilidade é ainda maior:

$$P(C|B) = \frac{n(C \cap B)}{n(B)} = \frac{1 + 0 + 733 + 4}{33 + 63 + 901 + 833} = \frac{738}{1830} = 0{,}4033$$

onde C é o número de infectados pelo sarampo no estado do Amazonas nos 11 anos, e B é o número de infectados pelo sarampo a nível nacional no período de 2016 a 2019.

Analisando-se individualmente estes estados pelos dados do Gráfico 14 e por suas perspectivas populacionais de acordo com a tabela 7, percebe-se que em São Paulo com 595 infectados e uma população absoluta de 45.919.049 habitantes, tem-se uma taxa de 1,3 casos registrados a cada 100.000 habitantes. Para o Amazonas com 738 infectados e uma população absoluta de 4.144.597 habitantes, a taxa de 17,8 casos a cada 100.000 habitantes é ainda mais preocupante. Estes dados apontam que em termos proporcionais, o estado do Amazonas tem um índice de transmissão do vírus na população que precisa ser reduzido para se evitar um surto da doença. De acordo com os dados da tabela 17, foram aplicadas 3.391.613 doses da vacina tríplice viral no Amazonas no período de 2009 a 2019. Em termos populacionais, tem-se uma taxa de 81.832 imunizações a cada 100.000 habitantes.

A tabela 16 mostra o total de casos registrados por região no período de 2009 a 2019. Das cinco regiões que formam o território nacional, o Norte e o Sudeste concentram a maior parte dos 2.286 casos, 71% dos registros.

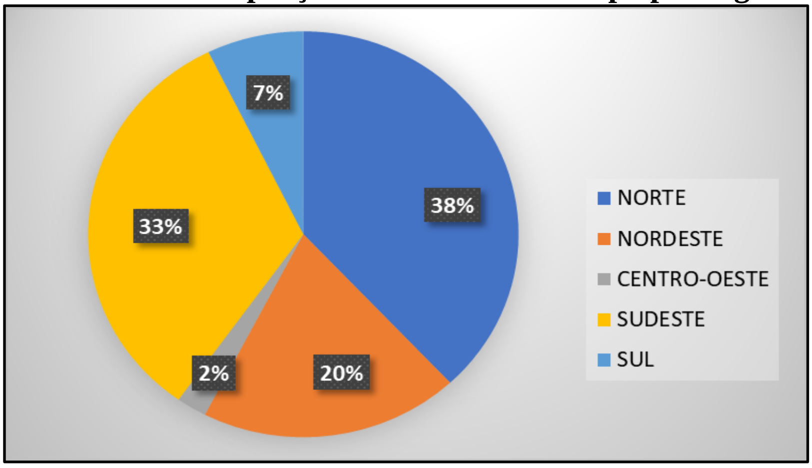
Tabela 16 – Internação hospitalar pelo sarampo por região

REGISTROS DE INTERNAÇÃO HOSPITALAR NO PERÍODO DE 2009 A 2019				
NORTE	NORDESTE	CENTRO-OESTE	SUDESTE	SUL
870	446	52	752	166
Total: 2.286				

Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

O Gráfico 15 faz uma distribuição proporcional do número de casos por região da tabela 16, onde percebe-se que o centro-oeste é o menos afetado com apenas 2% dos casos de infecções no período. Já a região norte, possui o maior número de infectados com um índice de 38%.

Gráfico 15 – Proporção dos casos de sarampo por região



Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Agora, vamos estudar cada uma das cinco regiões, separadamente, analisando os casos de infecção pelo sarampo registrados na tabela 16 e a população absoluta regional constante na tabela 8:

- **NORTE** – $\frac{870}{18.430.980} \approx \frac{4,7}{100.000}$ onde a taxa é de aproximadamente 4,7 pessoas infectadas a cada 100.000 habitantes;
- **NORDESTE** – $\frac{446}{57.071.654} \approx \frac{0,8}{100.000}$ onde a taxa é de aproximadamente 0,8 pessoa infectada a cada 100.000 habitantes;
- **CENTRO-OESTE** – $\frac{52}{16.297.074} \approx \frac{0,3}{100.000}$ onde a taxa é de aproximadamente 0,3 pessoa infectada a cada 100.000 habitantes;
- **SUDESTE** – $\frac{752}{88.371.433} \approx \frac{0,9}{100.000}$ onde a taxa é de aproximadamente 0,9 pessoa infectada a cada 100.000 habitantes;
- **SUL** – $\frac{166}{29.975.984} \approx \frac{0,6}{100.000}$ onde a taxa é de aproximadamente 0,6 pessoa infectada a cada 100.000 habitantes.

Este estudo mostra que a taxa de contaminação na região norte é cinco vezes maior que a taxa do sudeste, que possui a maior população absoluta entre as cinco regiões. Aqui, aponta-se a necessidade de um estudo para entender a causa de uma taxa de contaminação tão alta para se evitar um surto da doença.

Para uma análise das imunizações realizadas no período de 2009 a 2019, extraiu-se do site do DATASUS a tabela 17 com as doses da tríplice viral aplicadas por unidade federativa. Pelos dados desta tabela, observa-se que a média de imunizações por ano no período é de 9.900.903 doses aplicadas. Em 2019, registrou-se 17.086.383 de imunizações, a maior quantidade nos 11 anos observados. Em termos percentuais, isto equivale a 8,13% da população de 210.147.125 brasileira naquele ano.

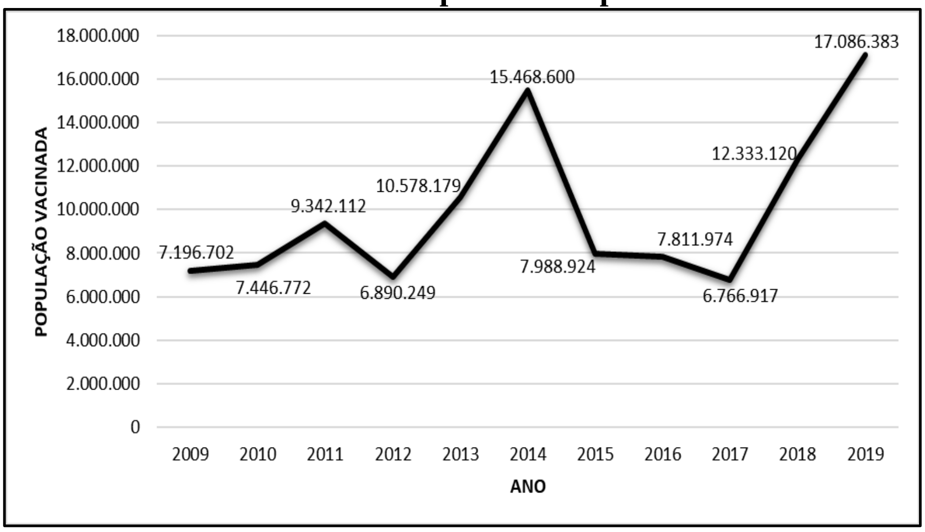
O Gráfico 16 mostra a variação no número de imunizações por ano. Nota-se que de 2014 a 2017, há uma queda de 56,2% na aplicação das vacinas, e posteriormente uma recuperação de 2017 a 2019 com um crescimento de 152,5%. Esta variação percentual expressiva aponta para a necessidade de uma investigação que determine sua causa.

Tabela 17 – Imunizações contra o vírus do sarampo de 2009 a 2019

DOSES DA TRÍPLICE VIRAL APLICADAS POR ESTADO E O DISTRITO FEDERAL												
	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	TOTAL
AC	39.888	37.407	57.707	28.328	37.575	69.120	24.113	20.725	22.676	52.195	38.400	428.134
AL	128.382	120.784	136.533	116.477	182.470	310.442	2.307.692	125.789	106.641	157.008	248.122	3.940.340
AP	31.735	30.372	34.509	32.854	41.969	75.259	23.932	27.116	25.649	52.514	41.789	417.698
AM	209.466	216.968	221.239	231.598	309.728	414.873	173.677	161.967	160.221	1.000.172	291.704	3.391.613
BA	476.765	491.001	605.635	495.178	848.253	1.115.690	323.145	431.937	437.965	689.719	1.085.636	7.000.924
CE	284.705	332.575	460.257	291.621	453.548	306.306	579.823	424.388	309.471	560.493	501.642	4.504.829
DF	123.003	115.170	156.025	85.990	161.083	214.930	61.419	192.236	94.634	160.197	217.111	1.581.798
ES	132.104	127.744	136.518	121.430	188.174	296.738	96.956	210.482	140.097	329.509	391.911	2.171.663
GO	202.210	203.059	274.164	215.621	319.035	504.930	164.708	191.776	183.392	263.824	334.528	2.857.247
MA	312.644	329.828	399.864	289.735	380.953	696.767	183.997	215.040	237.175	341.346	447.565	3.834.914
MT	104.955	114.639	129.186	115.685	183.423	284.460	89.086	106.581	104.569	193.307	199.486	1.625.377
MS	129.852	110.542	126.013	115.049	133.948	269.279	94.238	98.473	91.958	155.581	171.504	1.496.437
MG	579.585	562.038	634.038	651.281	958.613	1.463.645	438.949	695.957	713.951	1.485.273	1.669.230	9.852.560
PA	427.355	334.164	417.989	334.958	380.769	738.683	174.894	185.839	202.434	354.988	590.702	4.142.775
PB	138.208	415.629	145.779	108.653	226.340	311.672	90.319	124.109	117.513	172.182	259.397	2.109.801
PR	352.257	351.624	389.295	322.864	512.372	842.042	251.623	286.480	316.770	837.807	900.005	5.363.139
PE	358.471	356.797	463.148	313.412	705.192	633.250	322.363	442.599	322.754	524.356	818.345	5.260.687
PI	100.045	100.810	125.895	112.470	190.532	215.976	56.500	109.609	81.891	212.682	231.743	1.538.153
RJ	456.485	443.481	557.012	480.475	774.276	1.131.342	480.088	612.975	458.828	661.174	786.541	6.842.677
RN	91.015	99.923	112.701	98.477	166.876	231.663	64.646	83.883	77.100	108.651	197.683	1.332.618
RS	277.224	287.836	329.683	268.153	453.140	714.783	195.016	326.360	254.261	428.843	482.999	4.018.298
RO	60.442	55.513	67.299	62.523	98.991	157.745	44.586	52.868	64.208	156.916	121.134	942.225
RR	39.597	30.201	37.918	32.213	36.084	59.843	49.615	42.071	33.323	197.341	104.612	662.818
SC	245.423	249.759	252.180	231.688	326.952	536.938	193.532	219.699	205.914	559.787	667.527	3.689.399
SP	1.711.560	1.754.751	2.901.200	1.592.097	2.271.196	3.579.896	1.403.791	2.292.049	1.877.151	2.462.547	6.039.856	27.886.094
SE	77.282	72.274	81.833	78.361	134.520	163.691	58.651	84.364	71.573	126.041	162.409	1.110.999
TO	106.044	101.883	88.492	63.058	102.167	128.637	41.565	46.602	54.798	88.667	84.802	906.715
TOTAL	7.196.702	7.446.772	9.342.112	6.890.249	10.578.179	15.468.600	7.988.924	7.811.974	6.766.917	12.333.120	17.086.383	
TOTAL DE IMUNIZADOS NO PERÍODO:					108.909.932							

Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Gráfico 16 – Doses da vacina tríplice viral aplicadas de 2009 a 2019



Fonte: Adaptado http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Observa-se pela tabela 17 que o estado de São Paulo se destaca pelo número de imunizações com 27.886.094 doses da vacina aplicadas nestes últimos onze anos. Em termos percentuais, isto corresponde a 25,6% das 108.909.932 imunizações ocorridas.

Considerando-se os 26 estados e o Distrito Federal, a média de imunizações por unidade federativa no período é de 4.033.701. Para uma análise de abrangência nacional, as 108.909.932 doses aplicadas quando comparadas com a população de 210.147.125 brasileiros, nos fornece um índice de 51.825 doses aplicadas a cada 100.000 habitantes. Este último dado mostra a necessidade de uma investigação para determinar a causa dessa baixa taxa de imunização da população.

6 PROPOSTA DE ATIVIDADE

Uma atividade bem elaborada seguindo uma lógica sequencial, que favoreça a evolução do conhecimento e explore as habilidades e competências adquiridas pelo educando durante a aula, cria condições que levam a um ambiente dinâmico onde as aprendizagens ganham sentido.

A técnica da modelagem aliada à metodologia de uma sequência didática, para o reforço do processo ensino-aprendizagem, apoia-se em procedimentos por indagação e investigação para descrever fenômenos, situações e fatos do dia a dia pelas ferramentas matemáticas. Para a Estatística, a indagação e a investigação são indissociáveis por uma ocorrer à medida da outra. O objetivo da investigação é trabalhar a interdisciplinaridade e a contextualização da aprendizagem pela modelagem de um conjunto de dados, transformando-o em informações úteis pela estatística descritiva e fundamentando argumentações por cálculos probabilísticos.

Nesta proposta de atividade aberta, deseja-se que o educando explore os conceitos matemáticos propostos pelo currículo da Educação Básica das aprendizagens essenciais acerca das ferramentas da estatística e da probabilidade com o emprego de tecnologias. Todo o trabalho descritivo (quadros, tabelas e gráficos), analítico e probabilístico é implementado com o uso do *software* Excel por ser um programa de fácil manipulação. Todavia, ressalta-se que esta proposta pode ser tranquilamente trabalhada com um *software* livre como o Openoffice, Libreoffice, *Software* R, dentre outros.

Acredita-se que esse tipo de abordagem enriquece a exploração do conteúdo tornando o processo ensino-aprendizagem mais atraente e consistente. Apresentaremos nossas contribuições originais que tem como principal objetivo trazer “o mundo real” para dentro de uma sala de aula onde o educando possa aplicar seus conhecimentos adquiridos.

Infelizmente, esta proposta não foi implementada devido a Pandemia da Covid 19 e o consequente distanciamento social com a interrupção das aulas presenciais, uma vez que toda a estruturação da atividade é pensada para o ensino presencial com uma ativa interação entre professor e alunos.

6.1 Estrutura da atividade

Esta atividade foi estruturada para ser implementada com estudantes da segunda ou terceira séries do Ensino Médio, por serem nestas etapas da Educação Básica que se trabalham os conteúdos de Estatística Descritiva e Probabilidade. Assim, o educando tem a oportunidade de colocar em prática o conhecimento que lhe é proposto dando sentido à aprendizagem com a expectativa de que ela se torne significativa. Pode-se perfeitamente trabalhar atividades como esta, em séries anteriores, empregando conhecimentos estatísticos que começam a ser desenvolvidos no sexto ano do Ensino Fundamental II.

Considerando o Ensino Médio, sugere-se que esta proposta de atividade seja aplicada depois que o docente lecionou todo o conteúdo de Estatística da série, colocando o aluno para exercitar os conhecimentos adquiridos.

Propõe-se que a atividade seja feita em oito horas-aula de cinquenta minutos, pela metodologia da sequência didática, em um trabalho de quatro etapas. Em uma situação ideal, cada aluno fará sua atividade individualmente utilizando um computador. Caso não seja possível, o docente pode criar grupos. Recomenda-se que, ao final das atividades, os educandos interajam entre si para uma troca de experiências discutindo os resultados encontrados.

6.1.1 Primeira etapa

Com uma hora-aula, promove-se a contextualização da aprendizagem onde os alunos recebem do professor as orientações acerca do objeto de pesquisa, sua estrutura e as condições coletivas de produção dos saberes envolvidos.

No intuito de instigar o educando, propõe-se uma atividade aberta com diversas possibilidades de encaminhamento para fins de modelagem, análises e inferências sobre o número de indivíduos infectados pelo vírus da dengue no Brasil. A importância do tema escolhido se deve à recorrência dessa doença no país, sua presença em todo o território nacional e o fato de ser considerada um dos principais problemas de saúde pública pelo Ministério da Saúde.

No intuito de delimitar a pesquisa, sugere-se uma amostra de dados que abrangem o período de 2009 a 2019, retirada do site do Ministério da Saúde - Sistema de Informações

Hospitalares do SUS (DATASUS)[12]. Para entender como se dá o acesso às informações extraídas do site DATASUS, observa-se a tabela 18, onde constam os dados referentes ao ano de 2009. As outras dez tabelas, que estão no Anexo II, tem a mesma configuração e cada aluno deverá ter todas salvas no computador.

Tabela 18 – Registros de infecção pelo vírus da dengue em 2009

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2009						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região C-Oeste	Total
11 Rondônia	3.608	-	-	-	-	3.608
12 Acre	640	-	-	-	-	640
13 Amazonas	227	-	-	-	-	227
14 Roraima	798	-	-	-	-	798
15 Pará	4.751	-	-	-	-	4.751
16 Amapá	146	-	-	-	-	146
17 Tocantins	913	-	-	-	-	913
21 Maranhão	-	2.156	-	-	-	2.156
22 Piauí	-	2.171	-	-	-	2.171
23 Ceará	-	2.068	-	-	-	2.068
24 Rio Grande do Norte	-	588	-	-	-	588
25 Paraíba	-	651	-	-	-	651
26 Pernambuco	-	537	-	-	-	537
27 Alagoas	-	363	-	-	-	363
28 Sergipe	-	292	-	-	-	292
29 Bahia	-	16.451	-	-	-	16.451
31 Minas Gerais	-	-	2.619	-	-	2.619
32 Espírito Santo	-	-	3.189	-	-	3.189
33 Rio de Janeiro	-	-	1.050	-	-	1.050
35 São Paulo	-	-	851	-	-	851
41 Paraná	-	-	-	152	-	152
42 Santa Catarina	-	-	-	7	-	7
43 Rio Grande do Sul	-	-	-	5	-	5
50 Mato Grosso do Sul	-	-	-	-	686	686
51 Mato Grosso	-	-	-	-	3.867	3.867
52 Goiás	-	-	-	-	2.669	2.669
53 Distrito Federal	-	-	-	-	133	133
00 Ignorado/exterior	-	-	-	-	-	-
Total	11.083	25.277	7.709	164	7.355	51.588
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Fonte: http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

[12] http://www2.datasus.gov.br/DATASUS/index.php?area=02

6.1.2 Segunda etapa

Com duas horas-aula, o professor atua com intervenções que visam auxiliar os alunos na aplicação das ferramentas matemáticas e no domínio de uso do *software* Excel, procurando adequar as ações com os objetivos da aprendizagem pela modelagem matemática, que será colocada em prática na quarta etapa. Neste momento, os alunos devem aprender a trabalhar com o *software* citado, construindo tabelas, quadros e gráficos. Para este fim, pode-se usar as informações da tabela 18. No Anexo I, apresenta-se um esquema ilustrativo que mostra alguns princípios básicos para iniciar este trabalho em sala.

6.1.3 Terceira etapa

Com duas horas-aula, promove-se uma oficina para que os alunos interajam entre si e busquem, por meios individuais e coletivos, as informações necessárias para fazer a modelagem matemática do objeto de estudo. O professor deve esclarecer as dúvidas sobre o conteúdo de Estatística, sobre o uso do Excel e gerenciar o processo de troca de experiências entre os alunos.

6.1.3 Quarta etapa

Com três horas-aula, o professor orienta seus alunos na elaboração do trabalho atuando para que estes coloquem em prática os conhecimentos adquiridos. Sugere-se que todos executem as lições junto com o docente, que conduzirá o processo de modelagem e inferências estatísticas.

De posse das onze tabelas, o professor inicia com os alunos a organização das informações que tenham certa similaridade e o estudo delas. Ressalta-se que o professor neste momento deve agir como mentor, estabelecendo o tipo de tabela e gráfico a serem produzidos e orientando seus alunos a usarem o conhecimento e a criatividade.

Após a organização das informações, o educando deverá ter a habilidade e o conhecimento para eliminar redundâncias e agrupar variáveis que tenham relação. A tabela 19, apresenta um exemplo de organização dos dados originais. As análises e inferências construídas a partir dela, são sugestões dentre inúmeras possibilidades de conduzir o trabalho com os educandos.

Ao finalizar o trabalho em sala com os discentes, o professor deverá pedir um relatório final, como tarefa, com mais algumas análises acerca da tabela 19. O objetivo é gerar um momento em que eles exercitem os conhecimentos sozinhos, demonstrem as aprendizagens e reforcem a compreensão dos conceitos estatísticos na prática.

6.2 Atividade: modelagem, análises e inferências a partir de um conjunto de dados

A seguir, apresenta-se uma sugestão de condução para a atividade.

Tabela 19 – Número de infecções pelo vírus da dengue de 2009 a 2019

MORBIDADE HOSPITALAR CID 10 - DENGUE													
ANO		**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**Total**
REGIÃO NORTE	AC	640	754	498	209	235	537	512	107	78	184	294	**4.048**
	AP	146	190	248	143	91	115	193	49	37	38	18	**1.268**
	AM	227	575	1.966	225	484	185	287	273	124	53	107	**4.506**
	PA	4.751	6.828	8.369	5.686	4.030	1.934	2.265	3.050	2.178	824	1.053	**40.968**
	RO	3.608	5.051	2.206	1.528	3.029	1.406	1.300	1.515	614	343	393	**20.993**
	RR	798	1.175	234	304	71	62	114	51	76	20	109	**3.014**
	TO	913	1.277	1.075	1.061	1.068	795	468	417	310	142	595	**8.121**
	Total	**11.083**	**15.850**	**14.596**	**9.156**	**9.008**	**5.034**	**5.139**	**5.462**	**3.417**	**1.604**	**2.569**	
REGIÃO NORDESTE	AL	363	2.742	828	1.254	514	608	1.181	890	174	155	1.336	**10.045**
	BA	16.451	10.628	9.127	10.424	9.452	3.168	7.937	8.966	1.855	1.070	3.723	**82.801**
	CE	2.068	5.181	8.993	3.923	3.340	3.207	6.331	3.234	1.742	628	1.647	**40.294**
	MA	2.156	3.261	5.723	4.162	2.832	2.690	3.884	5.189	2.261	548	1.981	**34.687**
	PB	651	2.194	2.442	1.307	1.742	1.013	893	2.417	393	1.517	1.235	**15.804**
	PE	537	4.325	3.713	2.892	775	656	1.475	3.194	468	1.109	1.843	**20.987**
	PI	2.171	3.252	2.752	2.365	1.500	1.406	1.222	775	578	246	750	**17.017**
	RN	588	1.017	2.654	1.544	1.540	1.183	536	2.280	180	813	637	**12.972**
	SE	292	220	652	539	132	162	204	135	36	20	1.067	**3.459**
	Total	**25.277**	**32.820**	**36.884**	**28.410**	**21.827**	**14.093**	**23.663**	**27.080**	**7.687**	**6.106**	**14.219**	
REGIÃO CENTRO OESTE	DF	133	786	376	195	833	545	410	762	265	83	1.065	**5.453**
	GO	2.669	9.872	2.956	1.660	6.589	5.592	9.034	3.794	3.006	4.329	5.710	**55.211**
	MT	3.867	4.467	981	2.794	1.838	513	904	738	370	387	718	**17.577**
	MS	686	3.900	857	585	2.914	375	2.211	2.448	302	313	2.895	**17.486**
	Total	**7.355**	**19.025**	**5.170**	**5.234**	**12.174**	**7.025**	**12.559**	**7.742**	**3.943**	**5.112**	**10.388**	
REGIÃO SUDESTE	ES	3.189	1.856	3.345	1.056	1.658	827	1.789	2.008	793	645	2.273	**19.439**
	MG	2.619	6.765	2.220	1.464	7.048	2.088	5.560	10.126	1.651	1.182	9.464	**50.187**
	RJ	1.050	4.099	8.943	3.558	4.035	494	2.505	2.225	306	380	552	**28.147**
	SP	851	5.981	3.205	1.138	4.101	4.134	14.289	4.680	834	919	10.238	**50.370**
	Total	**7.709**	**18.701**	**17.713**	**7.216**	**16.842**	**7.543**	**24.143**	**19.039**	**3.584**	**3.126**	**22.527**	
REGIÃO SUL	PR	152	2.012	1.785	344	3.261	1.848	2.779	2.615	342	209	1.622	**16.969**
	RS	5	110	58	18	47	31	177	225	41	15	80	**807**
	SC	7	36	23	14	29	14	108	431	45	25	131	**863**
	Total	**164**	**2.158**	**1.866**	**376**	**3.337**	**1.893**	**3.064**	**3.271**	**428**	**249**	**1.833**	
Total por ano		**51.588**	**88.554**	**76.229**	**50.392**	**63.188**	**35.588**	**68.568**	**62.594**	**19.059**	**16.197**	**51.536**	
TOTAL DE CASOS REGISTROS NO PERÍODO:						**583.493**							

Fonte: Adaptada http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Pela tabela 19, a menor incidência no número de casos por infecção do vírus da dengue foi em 2018, com 16.197 registros e o ano com o maior número absoluto foi 2010, com 88.554. Observa-se que no período, a média de indivíduos infectados por ano é de 53.045.

Analisando-se o número de infectados nos 26 estados e o Distrito Federal, a média é de 21.611 registros por unidade federativa. Ao calcular o desvio médio da distribuição de infectados pelas 27 unidades da federação, obtemos o valor de 15.539 registros. O desvio médio é o erro cometido ao substituir cada valor pela sua média. Nota-se que os valores da média e do desvio médio estão distantes com uma diferença de 6.072 casos de infecção. Isto mostra que enquanto algumas unidades federativas têm valores baixos de infecção, como o estado do Rio Grande do Sul com 807 casos registrados no período amostral, o estado da Bahia tem mais que 102 vezes este número, com 82.801 indivíduos infectados. Atenta-se para uma preocupação com este dado que não deve ficar somente em número pois, ao calcularmos a probabilidade de um indivíduo infectado no período de 2009 a 2019 ser do estado da Bahia é de

$$\mathrm{P(A)} = \frac{82.801}{583.493} = 0{,}1419$$

onde A é o número de infectados no estado da Bahia, em relação à população de infectados a nível nacional no período. Uma probabilidade alta ao se levar em consideração que o estudo abrange 26 estados e o Distrito Federal.

Ainda analisando este último dado, considere uma amostra de 100 indivíduos escolhidos aleatoriamente entre os 583.493 infectados no período estudado. Qual seria a probabilidade de termos exatamente 10 indivíduos provenientes do estado da Bahia? Pelo cálculo da distribuição binomial das probabilidades temos

$$\mathrm{P}(10) = \frac{100!}{(100-10)!\ 10!}\ .\ 0{,}1419^{10}\ .\ (1-0{,}1419)^{100-10}$$

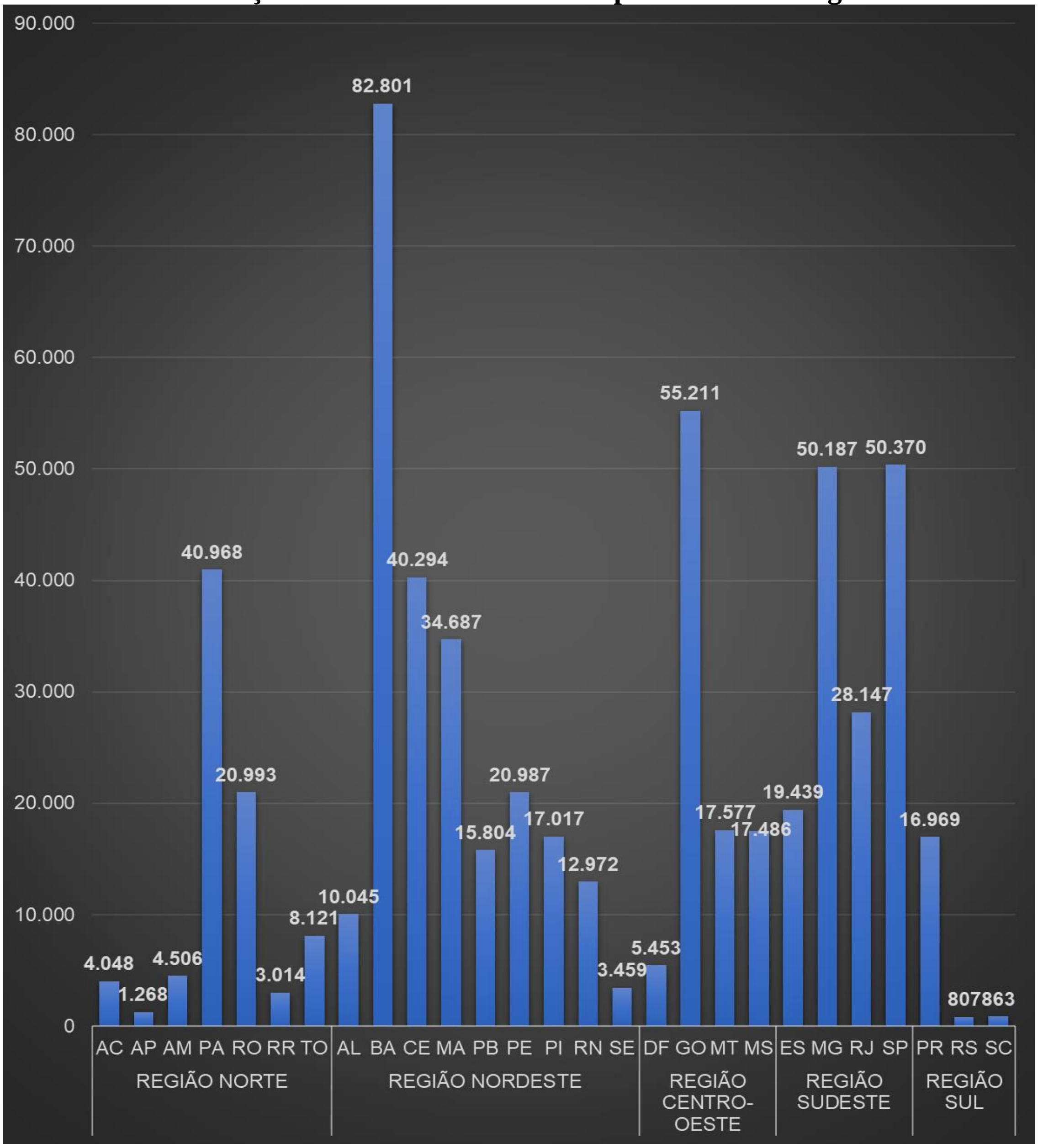
$$\mathrm{P}(10) = \frac{100!}{90!\ 10!}\ .\ 0{,}1419^{10}\ .\ 0{,}8581^{90}$$

$$\mathrm{P}(10) = 0{,}05977$$

onde $\mathrm{P}(10)$ é a probabilidade de se encontrar exatamente 10 indivíduos infectados pelo vírus da Dengue em uma amostra de 100. Ou seja, levando em consideração a aleatoriedade para se escolher uma amostra, tem-se aproximadamente 6% de chance para que 10% de uma amostra seja influenciada por indivíduos infectados provenientes do estado da Bahia.

Para se ter uma compreensão visual da tabela 19, elabora-se o Gráfico 17 com a distribuição do número de infectados por unidade da federação e região.

Gráfico 17 – Distribuição do número de infectados pelo vírus da dengue de 2009 a 2019



Fonte: Adaptada http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Outro dado que chama a atenção no Gráfico 17 são os registros da doença na região sul com 18.639 casos, o menor entre as cinco regiões. Os estados que possuem a menor quantidade de infectados em ordem crescente são: Rio Grande do Sul, Santa Catarina e o Amapá. Nota-se que o Distrito Federal tem 46,12% de casos registrados (5.453) a mais que os estados do Amapá, Rio Grande do Sul e Santa Catarina juntos (2.938). Ao se estudar os dados por região nos onze anos observados, temos a necessidade de construir a quadro 2 com uma distribuição dos infectados.

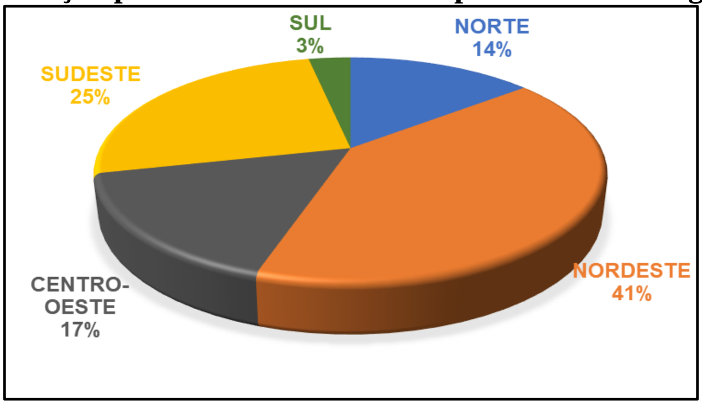
Quadro 2 – Número de infecções pelo vírus da dengue registradas por região

PERÍODO DE 2009 A 2019				
NORTE	NORDESTE	CENTRO-OESTE	SUDESTE	SUL
82.918	238.066	95.727	148.143	18.639
Total de casos no período:	583.493			

Fonte: Adaptada http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Neste quadro, observa-se que a região nordeste apresentou a maior incidência do número de casos que corresponde a aproximadamente 41% dos registros, enquanto a região sul teve a menor, com aproximadamente 3%. Isso inclusive, destaca a relação entre o quadro 2 e o Gráfico 18.

Gráfico 18 – Distribuição percentual de infectados pelo vírus da dengue de 2009 a 2019



Fonte: Adaptada http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Como mencionado anteriormente, a região nordeste chama a atenção quando comparada às outras regiões, com uma incidência discrepante no número de casos. A tabela 20 mostra que 34,78% dos infectados são do estado da Bahia e 1,45% de Sergipe.

Tabela 20 – Registros de infecção pelo vírus da dengue no Nordeste de 2009 a 2019

ESTADO	NÚMERO DE CASOS	PERCENTUAL
AL	10.045	4,22%
BA	82.801	34,78%
CE	40.294	16,93%
MA	34.687	14,57%
PB	15.804	6,64%
PE	20.987	8,82%
PI	17.017	7,15%
RN	12.972	5,45%
SE	3.459	1,45%

Fonte: Adaptada http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

Nota-se pela tabela 20 e pelo Gráfico 18 que a baixa incidência de casos no estado de Sergipe contrasta com o alto índice do Nordeste. Analisando-se o quadro 3, temos uma média de 314 infecções por ano no período. No entanto, o ano de 2019 aponta para um surto da doença com o registro de 1.067 casos, que é 3,4 vezes maior que a média mencionada nos onze anos analisados. Este dado chama a atenção para a necessidade de uma investigação para se determinar sua causa.

Quadro 3 – Infecções pelo vírus da dengue no estado de Sergipe

NÚMERO DE CASOS REGISTRADOS NO ESTADO DE SERGIPE NO PERÍODO DE ONZE ANOS										
2009	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
292	220	652	539	132	162	204	135	36	20	1.067
MÉDIA DE CASOS REGISTRADOS POR ANO:				314						

Fonte: Adaptada http://www2.datasus.gov.br/DATASUS/index.php?area=02
Acesso em: 10 de setembro de 2020.

7 CONSIDERAÇÕES FINAIS

O ensino da Matemática permeia todos os níveis acadêmicos, desde a formação básica à pós-graduação. O principal objetivo deste trabalho dissertativo foi contribuir para um ensino de qualidade e moderno de forma que o educando compreenda e aplique corretamente os conceitos da Estatística Descritiva e da Probabilidade. Ousar no ensino da desses conteúdos contribui para enriquecer o aprendizado pelo exercício criativo da intuição e pelo raciocínio lógico dedutivo, motiva a busca do conhecimento e do interesse por esta disciplina.

Desenvolver atividades investigativas nas aulas de matemática, coloca o educando em posição de protagonista na aquisição de seus saberes. Não se trata de eximir o professor da sua tarefa de ensinar. O educador é parte integrante de todo o processo de ensino e aprendizagem. É o condutor de uma metodologia voltada para fins educacionais pela proposta de uma atividade que dá sentido às aprendizagens pela sequência didática. Neste ponto, o professor coloca para os alunos uma situação problema relacionada a um tema cotidiano, estabelece os objetivos da investigação, indica as fontes para a pesquisa e coleta de dados, discute os conceitos estatísticos a serem trabalhados e orienta na elaboração das conclusões e sínteses. Aqui, procura-se favorecer a aplicação dos conhecimentos tornando a sala de aula um ambiente de agitação intelectual.

Pelo estudo de três bancos de dados sobre o censo demográfico, imunizações contra o sarampo e o número de infectados pelo vírus do sarampo, buscou-se mostrar a importância da Estatística e da Probabilidade para investigar, modelar por gráficos e tabelas, argumentar com base em cálculos probabilísticos e descrever o comportamento dos dados por análises sequenciais. Em seguida, elaborou-se uma proposta de atividade pela modelagem de um banco de dados sobre o número de infectados pelo vírus da dengue que pode ser desenvolvida com alunos da educação básica. Nesta atividade procura-se transformar um grande volume de dados em um material organizado, informativo e útil para análises sob diferentes aspectos. Estas modelagens foram implementadas no *software* Excel, mas também podem ser trabalhadas com qualquer *software* que permite construir tabelas e gráficos. Devido a Pandemia da Covid 19 e a consequente interrupção das aulas presenciais, não foi possível aplicar a proposta de atividade e aprofundar na pesquisa. Contudo, ao final do desenvolvimento deste estudo, reitera-se que é possível promover a aprendizagem significativa das definições da estatística e da teoria da probabilidade, pela técnica da modelagem apoiada pela metodologia da sequência didática.

Entende-se que a metodologia da sequência didática, é uma alternativa pedagógica para o ensino e a aprendizagem da matemática que coloca os alunos diante de situações-problema que, embora tenham interesse em resolver, não possuem, necessariamente, de antemão, ideias e ferramentas para isso. Pela modelagem de um dado real, pode-se pautar pela realidade e desenvolver conceitos pela interdisciplinaridade ou transdisciplinaridade, esclarecer ideias e sugerir caminhos distintos, ajudando a constituir um olhar mais questionador.

Nesse sentido, uma atividade desenvolvida em termos da sequência didática e que faça uso das definições da estatística descritiva e dos teoremas da probabilidade, caracteriza-se pelo caráter investigativo bem como pela possibilidade de fazer emergir conhecimentos que os alunos já possuem, permitindo aos mesmos reelaborá-los, ou fazer surgir a necessidade de construção de novos conhecimentos matemáticos.

Durante as pesquisas para a elaboração deste trabalho, percebe-se que há um vasto campo no qual os profissionais da educação possam estimular ou provocar o interesse de seu aluno pela Matemática. São vários os desafios para acompanhar essa Geração Z que nasceu em um berço impregnado de novas tecnologias. Além disso, muitas questões ficam em aberto para futuras pesquisas:

- A neurociência cognitiva ou neuropsicologia estuda as capacidades relativas à inteligência como a linguagem, a memória, a atenção, ao aprendizado entre outras. Tendo em vista esse conceito, como está o avanço dessa ciência e de qual forma podemos empregá-la no processo ensino-aprendizagem da Estatística para que nossos alunos consigam uma melhor depreensão desta ciência?
- As metodologias ativas e as tecnologias estão à disposição dos docentes. Como escolher a melhor ferramenta para auxiliá-lo no ensino da Estatística?

Ao procurar novos estudos e aprofundar nas metodologias da aprendizagem, obtém-se novos conhecimentos e técnicas que, porventura, geram novas práticas e conclusões.

Espera-se que as ideias aqui expostas, contribuam como fonte de apoio e incentivo para professores da educação básica. O intuito é criar um alicerce que possa assegurar ao educando a oportunidade de aprender os conceitos da Estatística e da Probabilidade, dada a importância dessas ferramentas matemáticas na formação dos futuros profissionais e cidadãos.

REFERÊNCIAS

ANDRADE, R; CALAPEZ, T; MELO, P; REIS, E. **Estatística aplicada** – Volume 1. 6ª ed, Edições Sílabo Ltda, 2015.

AUSUBEL, D. P. **Aquisição e retenção de conhecimentos: uma perspectiva cognitiva**. Lisboa: Plátano Edições Técnicas, 2003. Tradução de The acquisition and retention of knowledge: a cognitive view. (2000). Kluwer Academic Publishers

BARBOSA, J. C. **Modelagem na educação matemática: contribuições para o debate teórico.** In: Reunião anual da anped, 24., 2001, Caxambu. Rio de Janeiro: ANPED, 2001.

BASSANEZI, R. C. **Ensino-aprendizagem com modelagem matemática**. Disponível em: https://www.researchgate.net/publication/256007243. Acesso em: 29 de maio de 2020.

BRASIL, Instituto Brasileiro de Geografia e Estatística (IBGE). **Censo demográfico**. Disponível em: https://www.ibge.gov.br/estatisticas/sociais/populacao/9662-censo-demografico-2010.html?=&t=o-que-e. Acesso em: 20 de setembro de 2020.

BRASIL, Instituto Brasileiro de Geografia e Estatística (IBGE). **Estimativas da população**. Disponível em: https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=o-que-ec. Acesso em: 20 de setembro de 2020.

BRASIL, Instituto Brasileiro de Geografia e Estatística (IBGE). Sistema IBGE de Recuperação Automática – SIDRA. **Estimativas de população – estimapop – banco de metadados.** Disponível em: https://sidra.ibge.gov.br/pesquisa/estimapop/tabelas. Acesso em: 10 de setembro de 2020.

BRASIL, Ministério da Educação (MEC). **Base nacional curricular comum (BNCC)**. Disponível em: http://portal.mec.gov.br/conselho-nacional-de-educacao/base-nacional-comum-curricular-bncc. Acesso em: 27 de maio de 2020.

BRASIL, Ministério da Educação (MEC). **Relatório do Brasil no pisa 2018.** Disponível em: https://www.gov.br/inep/pt-br/areas-de-atuacao/avaliacao-e-exames-educacionais/pisa/resultados. Acesso em: 05 de dezembro de 2019.

BRASIL, Ministério da Saúde. **Manual de normas e procedimentos para vacinação.** Disponível em: http://bvsms.saude.gov.br/bvs/publicacoes/manual_procedimentos_vacinacao.pdf. Acesso em: 20 de setembro de 2020.

BRASIL, Ministério da Saúde. **Ministério da saúde atualiza casos de sarampo**. Disponível em: https://www.saude.gov.br/noticias/agencia-saude/45089-ministerio-da-saude-atualiza-casos-de-sarampo-19. Acesso em: 19 de setembro de 2020.

BRASIL, Ministério da Saúde. Portal da Saúde – DATASUS. **Informações de saúde (tabnet).** Disponível em: http://www2.datasus.gov.br/DATASUS/index.php?area=02. Acesso em: 10 de setembro de 2020.

BURAK, D.; PENTEADO, D.R. **As práticas que envolvem modelagem matemática na educação básica do paraná: uma meta-análise do epmem.** Disponível em: https://periodicos.ufn.edu.br/index.php/VIDYA/article/view/2646. Acesso em: 14 de janeiro de 2021.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica.** 6ª ed, Editora Saraiva, 2010.

CABRAL, N. F. **Sequências didáticas: estrutura e elaboração.** 1ª ed, Belém: SBEM / SBEM-PA, 2017.

CAMARGO, F.; DAROS, T. **A sala de aula inovadora: estratégias pedagógicas para fomentar o aprendizado ativo.** Porto Alegre: Penso Editora Ltda, 2018.

D'AMBROSIO, U. **Educação matemática: da teoria à prática.** 17ª ed, Campinas. São Paulo: Papirus Editora, 2009.

DUARTE, R. L. **Introdução à estatística e probabilidade: uma abordagem contextualizada no cotidiano dos alunos.** Disponível em: http://www.repositorio.ufc.br/bitstream/riufc/6158/1/2013_dis_rlduarte.pdf Acesso em: 14 de janeiro de 2021.

FARBER, B.; LARSON, R. **Estatística aplicada**. Tradução José Fernando Pereira Gonçalves. Revisão técnica Manoel Henrique Salgado. São Paulo; Pearson Education do Brasil, 2015.

FARIAS, A.M.L. **Estatística descritiva**. Disponível em: http://www.professores.uff.br/anafarias/. Acesso em: 24 de março de 2020.

FRANÇA, Organização para Cooperação e Desenvolvimento Econômico (OCDE). **Banco de dados pisa.** Disponível em: https://www.oecd.org/pisa/data/. Acesso em: 05 de dezembro de 2019.

FRANÇA, Organização para Cooperação e Desenvolvimento Econômico (OCDE). **Programme for international student assessment (pisa) results from pisa 2012.** Disponível em: https://www.oecd.org/pisa/keyfindings/pisa-2012-results.htm. Acesso em: 05 de dezembro de 2019.

LOPES, C.E. **O Ensino da estatística e da probabilidade na educação básica e a formação dos professores.** Cad. Cedes, Campinas, vol. 28, n. 74, p. 57-73, jan./abr. 2008.

LOPES, C.E.; MENDONÇA, L.O.; SOUZA, L.O.; SOUZA, A.C. **O ensino de estatística e probabilidade na educação básica: atividades e projetos gerados a partir de pesquisas de mestrado profissional.** Disponível em: https://periodicos.ufn.edu.br/index.php/VIDYA/article/view/245. Acesso em: 14 de janeiro de 2021.

MOREIRA, M. A. **O que é afinal aprendizagem significativa?** Revista Cultural La Laguna Espanha, 2012. Disponível em: http://moreira.if.ufrgs.br/oqueeafinal.pdf. Acesso em: 17 de agosto de 2020.

RIOS, E. M. **Estatística descritiva, probabilidade e estimação: noções para o ensino básico.** Disponível em: http://repositorio.bc.ufg.br/tede/handle/tede/4365. Acesso em: 11 de janeiro de 2021.

SILVA, R. T. **Interpretando dados do cotidiano: o ensino de estatística na educação básica.** Disponível em: http://www.bdtd.uerj.br/tde_busca/arquivo.php?codArquivo=8609. Acesso em: 11 de janeiro de 2021.

SOUZA, N. G. **Estatística e probabilidade no ensino médio.** Disponível em: http://www.locus.ufv.br/handle/123456789/24563. Acesso em: 11 de janeiro de 2021.

STEVANIM, L. F. **Sarampo de volta ao mapa.** ago. 2018. Fundação Oswaldo Cruz. Disponível em: https://portal.fiocruz.br/noticia/sarampo-de-volta-ao-mapa. Acesso em: 19 de setembro de 2020.

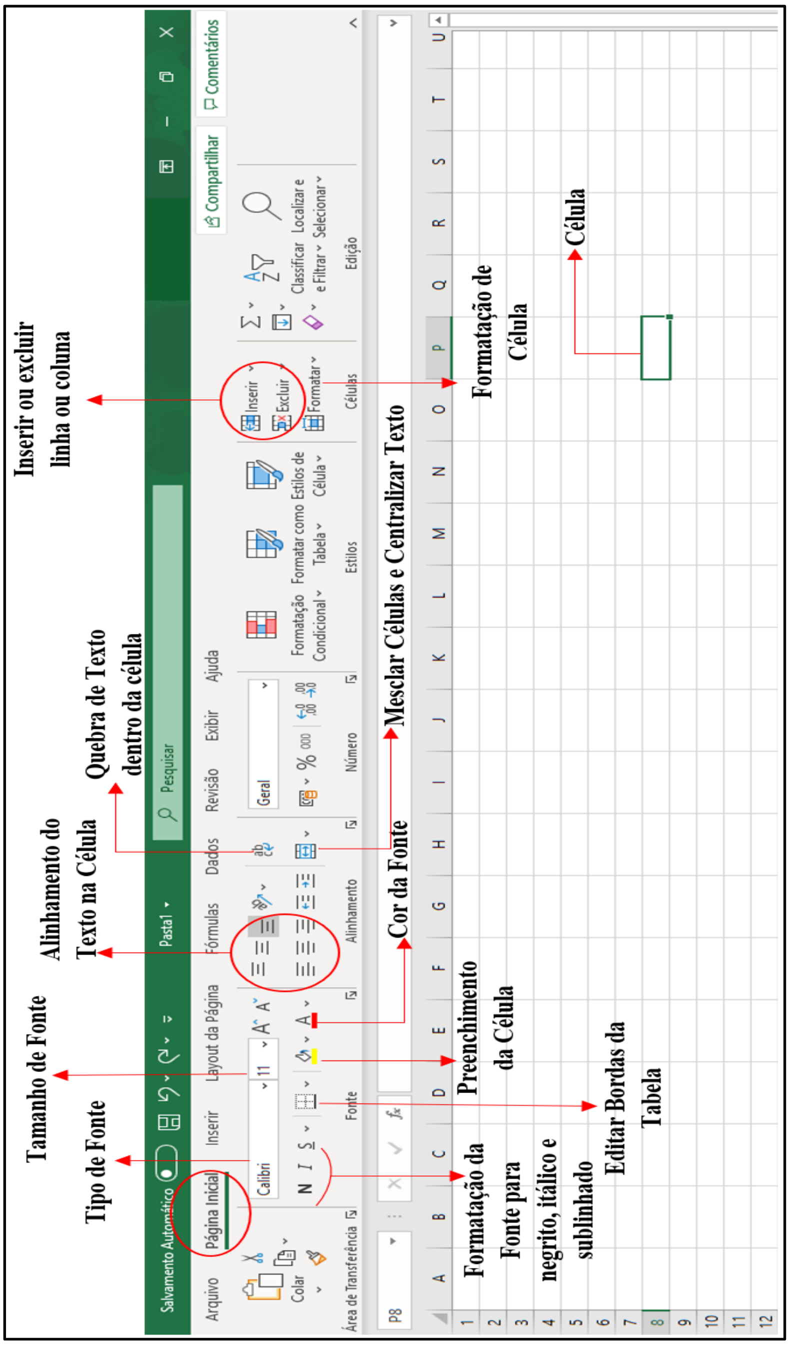
TRIOLA, M. F. **Introdução à estatística.** Tradução: Vera Regina Lima de Farias Flores. 10ª ed, Editora LTC, 2008.

ZABALA, A. **A prática educativa: como ensinar.** Tradução: Ernani F. da F. Rosa. Porto Alegre: ArtMed Editora SA, 1998.

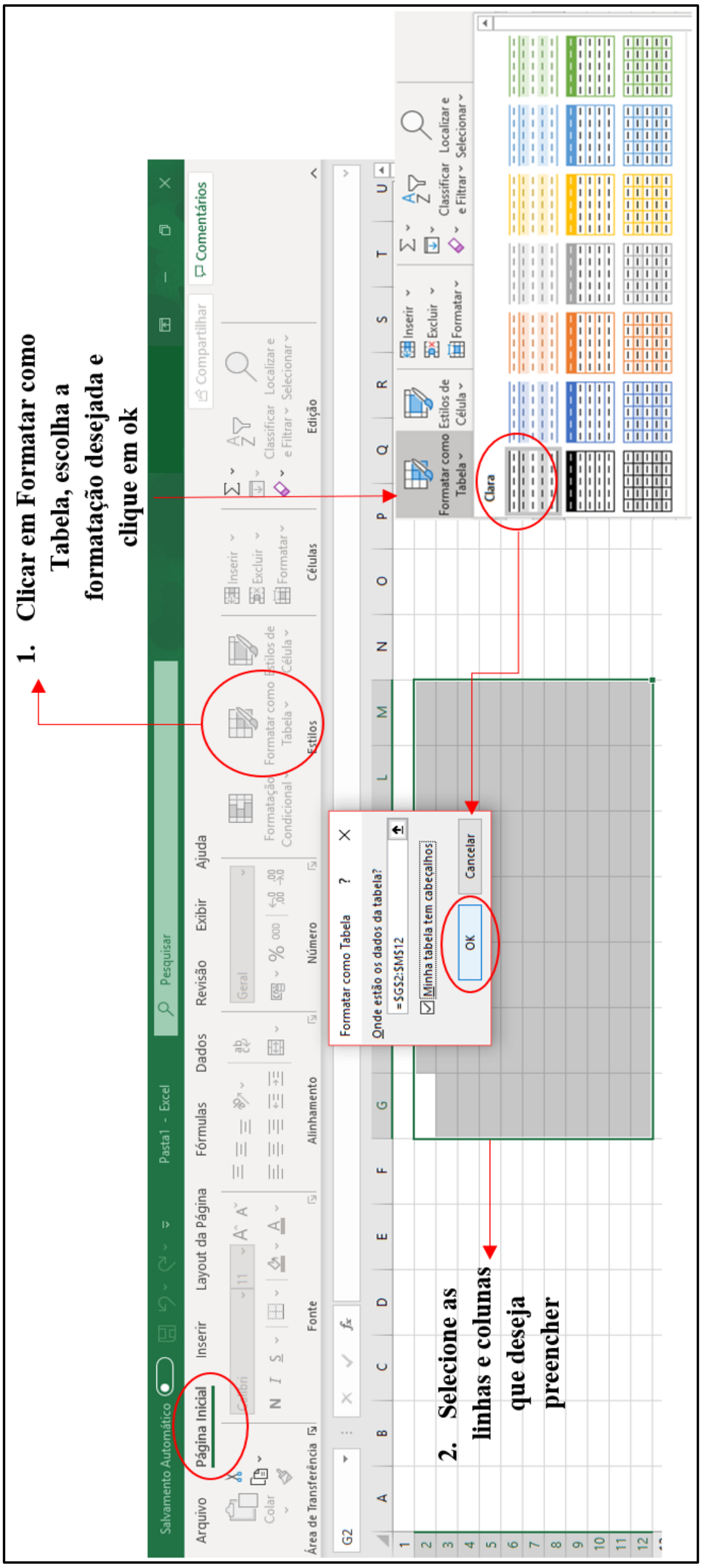
ANEXO I

Nesse anexo encontram-se algumas orientações para iniciar a elaboração de quadros, tabelas e gráficos no *software* Excel com os alunos.

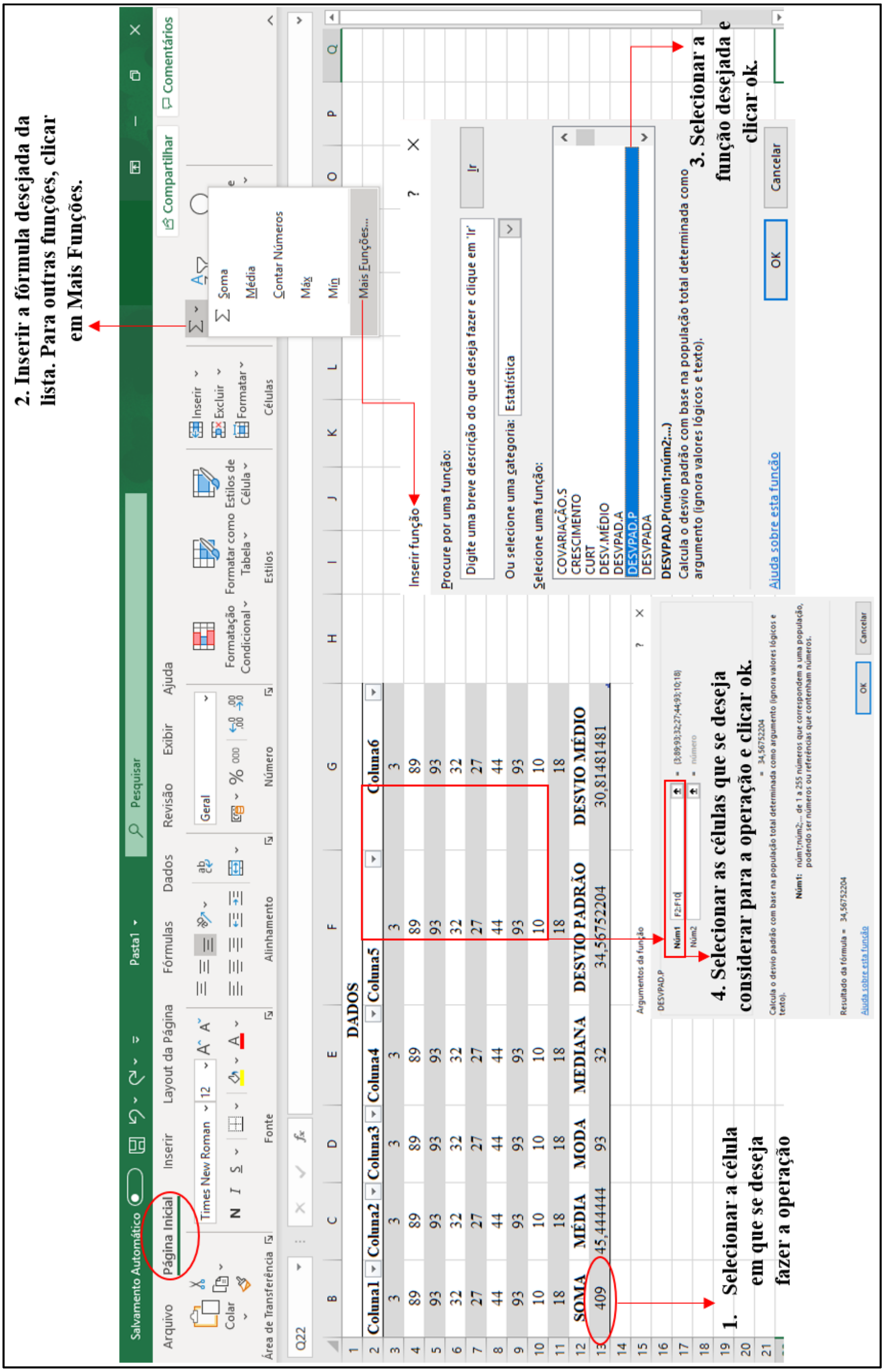
1 – Como fazer formatações gerais no Excel (aba página inicial):



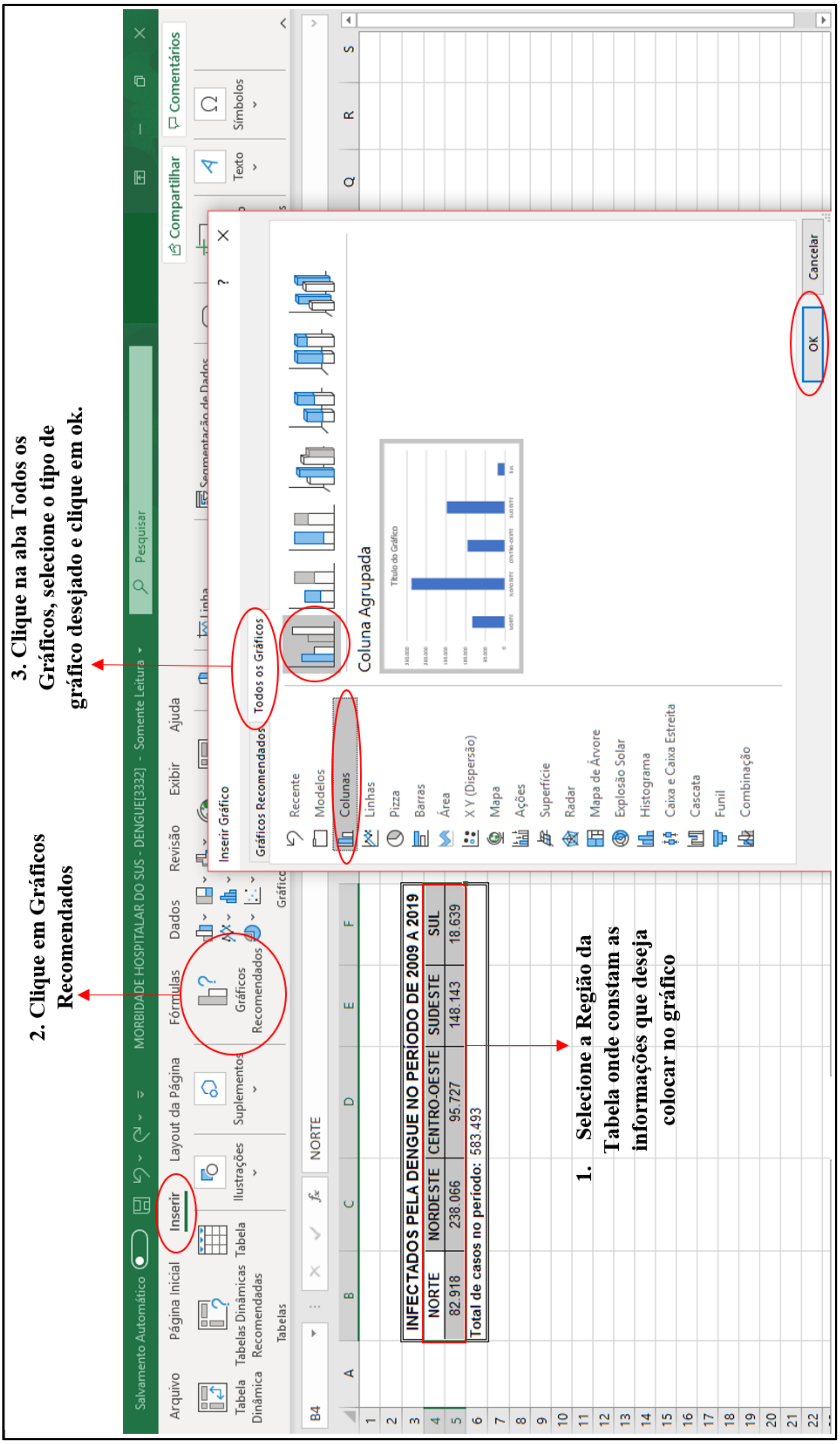
2 – Como criar uma tabela no Excel (aba página inicial):



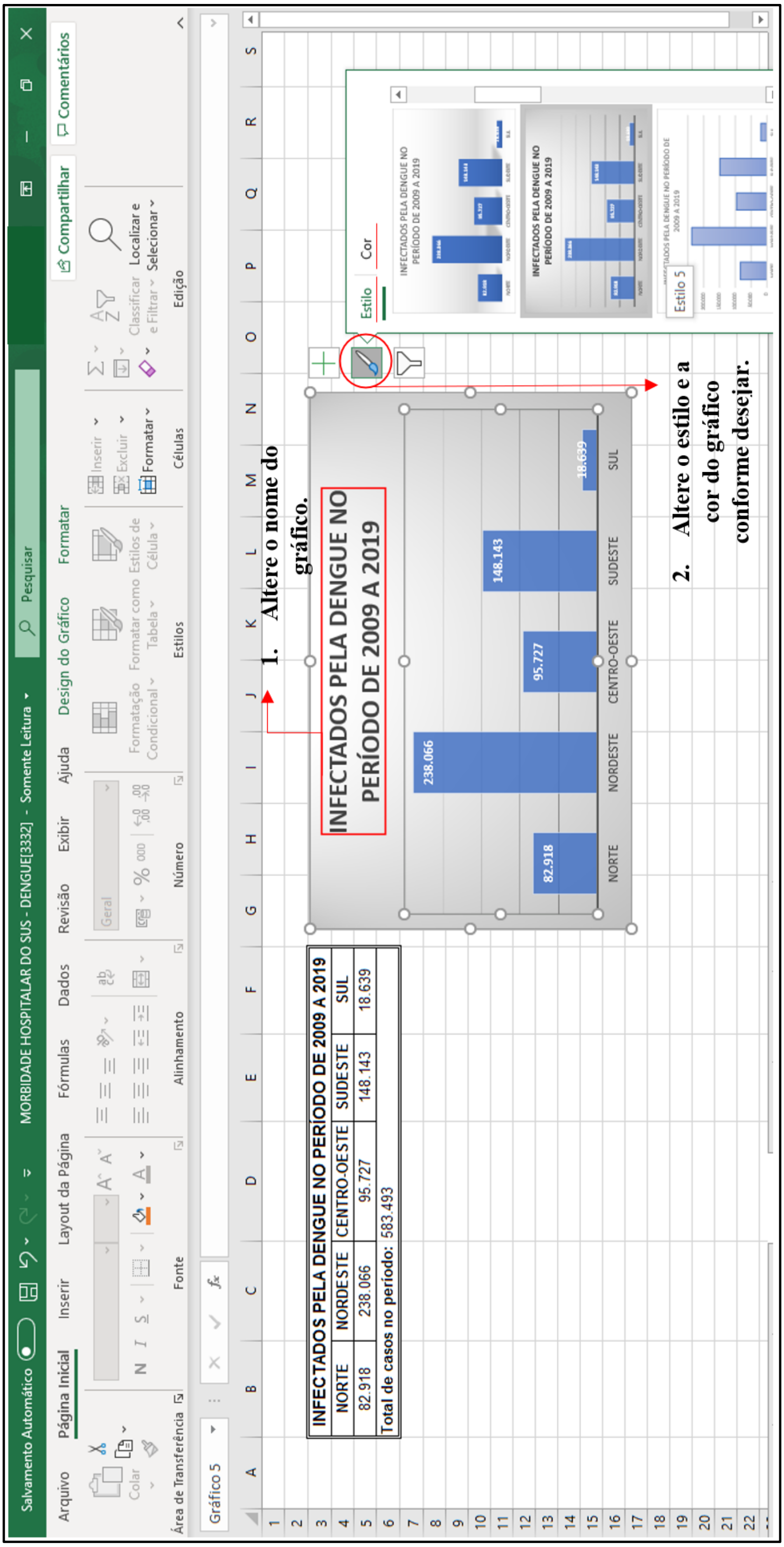
3 – Como inserir fórmulas no Excel (aba página inicial):



4 – Como fazer um gráfico no Excel (aba página inicial):



5 – Como fazer formatação de gráfico no Excel (aba página inicial):



ANEXO II

Nesse anexo encontram-se 10 tabelas com dados referentes ao número de infectados pelo vírus da dengue no Brasil, retirados do site do DATASUS e usados para fazer a modelagem da proposta de atividade do capítulo 6.

Registros de infecção pelo vírus da dengue em 2010

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2010						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	5051	-	-	-	-	5051
12 Acre	754	-	-	-	-	754
13 Amazonas	575	-	-	-	-	575
14 Roraima	1175	-	-	-	-	1175
15 Pará	6828	-	-	-	-	6828
16 Amapá	190	-	-	-	-	190
17 Tocantins	1277	-	-	-	-	1277
21 Maranhão	-	3261	-	-	-	3261
22 Piauí	-	3252	-	-	-	3252
23 Ceará	-	5181	-	-	-	5181
24 Rio Grande do Norte	-	1017	-	-	-	1017
25 Paraíba	-	2194	-	-	-	2194
26 Pernambuco	-	4325	-	-	-	4325
27 Alagoas	-	2742	-	-	-	2742
28 Sergipe	-	220	-	-	-	220
29 Bahia	-	10628	-	-	-	10628
31 Minas Gerais	-	-	6765	-	-	6765
32 Espírito Santo	-	-	1856	-	-	1856
33 Rio de Janeiro	-	-	4099	-	-	4099
35 São Paulo	-	-	5981	-	-	5981
41 Paraná	-	-	-	2012	-	2012
42 Santa Catarina	-	-	-	36	-	36
43 Rio Grande do Sul	-	-	-	110	-	110
50 Mato Grosso do Sul	-	-	-	-	3900	3900
51 Mato Grosso	-	-	-	-	4467	4467
52 Goiás	-	-	-	-	9872	9872
53 Distrito Federal	-	-	-	-	786	786
00 Ignorado/exterior	-	-	-	-	-	-
Total	15850	32820	18701	2158	19025	88554
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2011

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2011						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	2206	-	-	-	-	2206
12 Acre	498	-	-	-	-	498
13 Amazonas	1966	-	-	-	-	1966
14 Roraima	234	-	-	-	-	234
15 Pará	8369	-	-	-	-	8369
16 Amapá	248	-	-	-	-	248
17 Tocantins	1075	-	-	-	-	1075
21 Maranhão	-	5723	-	-	-	5723
22 Piauí	-	2752	-	-	-	2752
23 Ceará	-	8993	-	-	-	8993
24 Rio Grande do Norte	-	2654	-	-	-	2654
25 Paraíba	-	2442	-	-	-	2442
26 Pernambuco	-	3713	-	-	-	3713
27 Alagoas	-	828	-	-	-	828
28 Sergipe	-	652	-	-	-	652
29 Bahia	-	9127	-	-	-	9127
31 Minas Gerais	-	-	2220	-	-	2220
32 Espírito Santo	-	-	3345	-	-	3345
33 Rio de Janeiro	-	-	8943	-	-	8943
35 São Paulo	-	-	3205	-	-	3205
41 Paraná	-	-	-	1785	-	1785
42 Santa Catarina	-	-	-	23	-	23
43 Rio Grande do Sul	-	-	-	58	-	58
50 Mato Grosso do Sul	-	-	-	-	857	857
51 Mato Grosso	-	-	-	-	981	981
52 Goiás	-	-	-	-	2956	2956
53 Distrito Federal	-	-	-	-	376	376
00 Ignorado/exterior	-	-	-	-	-	-
Total	14596	36884	17713	1866	5170	76229
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2012

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2012						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	1528	-	-	-	-	1528
12 Acre	209	-	-	-	-	209
13 Amazonas	225	-	-	-	-	225
14 Roraima	304	-	-	-	-	304
15 Pará	5686	-	-	-	-	5686
16 Amapá	143	-	-	-	-	143
17 Tocantins	1061	-	-	-	-	1061
21 Maranhão	-	4162	-	-	-	4162
22 Piauí	-	2365	-	-	-	2365
23 Ceará	-	3923	-	-	-	3923
24 Rio Grande do Norte	-	1544	-	-	-	1544
25 Paraíba	-	1307	-	-	-	1307
26 Pernambuco	-	2892	-	-	-	2892
27 Alagoas	-	1254	-	-	-	1254
28 Sergipe	-	539	-	-	-	539
29 Bahia	-	10424	-	-	-	10424
31 Minas Gerais	-	-	1464	-	-	1464
32 Espírito Santo	-	-	1056	-	-	1056
33 Rio de Janeiro	-	-	3558	-	-	3558
35 São Paulo	-	-	1138	-	-	1138
41 Paraná	-	-	-	344	-	344
42 Santa Catarina	-	-	-	14	-	14
43 Rio Grande do Sul	-	-	-	18	-	18
50 Mato Grosso do Sul	-	-	-	-	585	585
51 Mato Grosso	-	-	-	-	2794	2794
52 Goiás	-	-	-	-	1660	1660
53 Distrito Federal	-	-	-	-	195	195
00 Ignorado/exterior	-	-	-	-	-	-
Total	9156	28410	7216	376	5234	50392
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2013

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2013						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	3029	-	-	-	-	3029
12 Acre	235	-	-	-	-	235
13 Amazonas	484	-	-	-	-	484
14 Roraima	71	-	-	-	-	71
15 Pará	4030	-	-	-	-	4030
16 Amapá	91	-	-	-	-	91
17 Tocantins	1068	-	-	-	-	1068
21 Maranhão	-	2832	-	-	-	2832
22 Piauí	-	1500	-	-	-	1500
23 Ceará	-	3340	-	-	-	3340
24 Rio Grande do Norte	-	1540	-	-	-	1540
25 Paraíba	-	1742	-	-	-	1742
26 Pernambuco	-	775	-	-	-	775
27 Alagoas	-	514	-	-	-	514
28 Sergipe	-	132	-	-	-	132
29 Bahia	-	9452	-	-	-	9452
31 Minas Gerais	-	-	7048	-	-	7048
32 Espírito Santo	-	-	1658	-	-	1658
33 Rio de Janeiro	-	-	4035	-	-	4035
35 São Paulo	-	-	4101	-	-	4101
41 Paraná	-	-	-	3261	-	3261
42 Santa Catarina	-	-	-	29	-	29
43 Rio Grande do Sul	-	-	-	47	-	47
50 Mato Grosso do Sul	-	-	-	-	2914	2914
51 Mato Grosso	-	-	-	-	1838	1838
52 Goiás	-	-	-	-	6589	6589
53 Distrito Federal	-	-	-	-	833	833
00 Ignorado/exterior	-	-	-	-	-	-
Total	9008	21827	16842	3337	12174	63188
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2014

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2014						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	1406	-	-	-	-	1406
12 Acre	537	-	-	-	-	537
13 Amazonas	185	-	-	-	-	185
14 Roraima	62	-	-	-	-	62
15 Pará	1934	-	-	-	-	1934
16 Amapá	115	-	-	-	-	115
17 Tocantins	795	-	-	-	-	795
21 Maranhão	-	2690	-	-	-	2690
22 Piauí	-	1406	-	-	-	1406
23 Ceará	-	3207	-	-	-	3207
24 Rio Grande do Norte	-	1183	-	-	-	1183
25 Paraíba	-	1013	-	-	-	1013
26 Pernambuco	-	656	-	-	-	656
27 Alagoas	-	608	-	-	-	608
28 Sergipe	-	162	-	-	-	162
29 Bahia	-	3168	-	-	-	3168
31 Minas Gerais	-	-	2088	-	-	2088
32 Espírito Santo	-	-	827	-	-	827
33 Rio de Janeiro	-	-	494	-	-	494
35 São Paulo	-	-	4134	-	-	4134
41 Paraná	-	-	-	1848	-	1848
42 Santa Catarina	-	-	-	14	-	14
43 Rio Grande do Sul	-	-	-	31	-	31
50 Mato Grosso do Sul	-	-	-	-	375	375
51 Mato Grosso	-	-	-	-	513	513
52 Goiás	-	-	-	-	5592	5592
53 Distrito Federal	-	-	-	-	545	545
00 Ignorado/exterior	-	-	-	-	-	-
Total	5034	14093	7543	1893	7025	35588
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2015

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2015						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	1300	-	-	-	-	1300
12 Acre	512	-	-	-	-	512
13 Amazonas	287	-	-	-	-	287
14 Roraima	114	-	-	-	-	114
15 Pará	2265	-	-	-	-	2265
16 Amapá	193	-	-	-	-	193
17 Tocantins	468	-	-	-	-	468
21 Maranhão	-	3884	-	-	-	3884
22 Piauí	-	1222	-	-	-	1222
23 Ceará	-	6331	-	-	-	6331
24 Rio Grande do Norte	-	536	-	-	-	536
25 Paraíba	-	893	-	-	-	893
26 Pernambuco	-	1475	-	-	-	1475
27 Alagoas	-	1181	-	-	-	1181
28 Sergipe	-	204	-	-	-	204
29 Bahia	-	7937	-	-	-	7937
31 Minas Gerais	-	-	5560	-	-	5560
32 Espírito Santo	-	-	1789	-	-	1789
33 Rio de Janeiro	-	-	2505	-	-	2505
35 São Paulo	-	-	14289	-	-	14289
41 Paraná	-	-	-	2779	-	2779
42 Santa Catarina	-	-	-	108	-	108
43 Rio Grande do Sul	-	-	-	177	-	177
50 Mato Grosso do Sul	-	-	-	-	2211	2211
51 Mato Grosso	-	-	-	-	904	904
52 Goiás	-	-	-	-	9034	9034
53 Distrito Federal	-	-	-	-	410	410
00 Ignorado/exterior	-	-	-	-	-	-
Total	5139	23663	24143	3064	12559	68568
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2016

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2016						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	1515	-	-	-	-	1515
12 Acre	107	-	-	-	-	107
13 Amazonas	273	-	-	-	-	273
14 Roraima	51	-	-	-	-	51
15 Pará	3050	-	-	-	-	3050
16 Amapá	49	-	-	-	-	49
17 Tocantins	417	-	-	-	-	417
21 Maranhão	-	5189	-	-	-	5189
22 Piauí	-	775	-	-	-	775
23 Ceará	-	3234	-	-	-	3234
24 Rio Grande do Norte	-	2280	-	-	-	2280
25 Paraíba	-	2417	-	-	-	2417
26 Pernambuco	-	3194	-	-	-	3194
27 Alagoas	-	890	-	-	-	890
28 Sergipe	-	135	-	-	-	135
29 Bahia	-	8966	-	-	-	8966
31 Minas Gerais	-	-	10126	-	-	10126
32 Espírito Santo	-	-	2008	-	-	2008
33 Rio de Janeiro	-	-	2225	-	-	2225
35 São Paulo	-	-	4680	-	-	4680
41 Paraná	-	-	-	2615	-	2615
42 Santa Catarina	-	-	-	431	-	431
43 Rio Grande do Sul	-	-	-	225	-	225
50 Mato Grosso do Sul	-	-	-	-	2448	2448
51 Mato Grosso	-	-	-	-	738	738
52 Goiás	-	-	-	-	3794	3794
53 Distrito Federal	-	-	-	-	762	762
00 Ignorado/exterior	-	-	-	-	-	-
Total	5462	27080	19039	3271	7742	62594
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2017

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2017						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	614	-	-	-	-	614
12 Acre	78	-	-	-	-	78
13 Amazonas	124	-	-	-	-	124
14 Roraima	76	-	-	-	-	76
15 Pará	2178	-	-	-	-	2178
16 Amapá	37	-	-	-	-	37
17 Tocantins	310	-	-	-	-	310
21 Maranhão	-	2261	-	-	-	2261
22 Piauí	-	578	-	-	-	578
23 Ceará	-	1742	-	-	-	1742
24 Rio Grande do Norte	-	180	-	-	-	180
25 Paraíba	-	393	-	-	-	393
26 Pernambuco	-	468	-	-	-	468
27 Alagoas	-	174	-	-	-	174
28 Sergipe	-	36	-	-	-	36
29 Bahia	-	1855	-	-	-	1855
31 Minas Gerais	-	-	1651	-	-	1651
32 Espírito Santo	-	-	793	-	-	793
33 Rio de Janeiro	-	-	306	-	-	306
35 São Paulo	-	-	834	-	-	834
41 Paraná	-	-	-	342	-	342
42 Santa Catarina	-	-	-	45	-	45
43 Rio Grande do Sul	-	-	-	41	-	41
50 Mato Grosso do Sul	-	-	-	-	302	302
51 Mato Grosso	-	-	-	-	370	370
52 Goiás	-	-	-	-	3006	3006
53 Distrito Federal	-	-	-	-	265	265
00 Ignorado/exterior	-	-	-	-	-	-
Total	3417	7687	3584	428	3943	19059
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2018

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2018						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	343	-	-	-	-	343
12 Acre	184	-	-	-	-	184
13 Amazonas	53	-	-	-	-	53
14 Roraima	20	-	-	-	-	20
15 Pará	824	-	-	-	-	824
16 Amapá	38	-	-	-	-	38
17 Tocantins	142	-	-	-	-	142
21 Maranhão	-	548	-	-	-	548
22 Piauí	-	246	-	-	-	246
23 Ceará	-	628	-	-	-	628
24 Rio Grande do Norte	-	813	-	-	-	813
25 Paraíba	-	1517	-	-	-	1517
26 Pernambuco	-	1109	-	-	-	1109
27 Alagoas	-	155	-	-	-	155
28 Sergipe	-	20	-	-	-	20
29 Bahia	-	1070	-	-	-	1070
31 Minas Gerais	-	-	1182	-	-	1182
32 Espírito Santo	-	-	645	-	-	645
33 Rio de Janeiro	-	-	380	-	-	380
35 São Paulo	-	-	919	-	-	919
41 Paraná	-	-	-	209	-	209
42 Santa Catarina	-	-	-	25	-	25
43 Rio Grande do Sul	-	-	-	15	-	15
50 Mato Grosso do Sul	-	-	-	-	313	313
51 Mato Grosso	-	-	-	-	387	387
52 Goiás	-	-	-	-	4329	4329
53 Distrito Federal	-	-	-	-	83	83
00 Ignorado/exterior	-	-	-	-	-	-
Total	1604	6106	3126	249	5112	16197
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						

Registros de infecção pelo vírus da dengue em 2019

Morbidade Hospitalar do SUS - por local de internação - Brasil						
Internações por Unidade da Federação e Região						
Lista Morb CID-10: Dengue [dengue clásssico]						
Período:2019						
Unidade da Federação	1 Região Norte	2 Região Nordeste	3 Região Sudeste	4 Região Sul	5 Região Centro-Oeste	Total
11 Rondônia	393	-	-	-	-	393
12 Acre	294	-	-	-	-	294
13 Amazonas	107	-	-	-	-	107
14 Roraima	109	-	-	-	-	109
15 Pará	1053	-	-	-	-	1053
16 Amapá	18	-	-	-	-	18
17 Tocantins	595	-	-	-	-	595
21 Maranhão	-	1981	-	-	-	1981
22 Piauí	-	750	-	-	-	750
23 Ceará	-	1647	-	-	-	1647
24 Rio Grande do Norte	-	637	-	-	-	637
25 Paraíba	-	1235	-	-	-	1235
26 Pernambuco	-	1843	-	-	-	1843
27 Alagoas	-	1336	-	-	-	1336
28 Sergipe	-	1067	-	-	-	1067
29 Bahia	-	3723	-	-	-	3723
31 Minas Gerais	-	-	9464	-	-	9464
32 Espírito Santo	-	-	2273	-	-	2273
33 Rio de Janeiro	-	-	552	-	-	552
35 São Paulo	-	-	10238	-	-	10238
41 Paraná	-	-	-	1622	-	1622
42 Santa Catarina	-	-	-	131	-	131
43 Rio Grande do Sul	-	-	-	80	-	80
50 Mato Grosso do Sul	-	-	-	-	2895	2895
51 Mato Grosso	-	-	-	-	718	718
52 Goiás	-	-	-	-	5710	5710
53 Distrito Federal	-	-	-	-	1065	1065
00 Ignorado/exterior	-	-	-	-	-	-
Total	2569	14219	22527	1833	10388	51536
Fonte: Ministério da Saúde - Sistema de Informações Hospitalares do SUS (SIH/SUS)						
Notas: Situação da base de dados nacional em 29/04/2016.						
Dados de janeiro de 2015 até março de 2016 sujeitos a retificação.						